Observation of the Higgs boson and measurement of its properties in the HWW* channel with the ATLAS detector at the LHC

Yichen Li

UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 517:
PARTICULES, NOYAUX ET COSMOS

Laboratoire: Laboratoire de l'Accélérateur Linéaire

THÈSE DE DOCTORAT
PHYSIQUE

par

Yichen LI

Observation du boson de Higgs et mesure de ses propriétés dans le canal $H \to WW^* \to \ell\nu\ell\nu$ avec le détecteur ATLAS au LHC

Date de soutenance:11/04/2015

Composition du jury:

Directeur de thèse:	Zhiqing Zhang	LAL
Co-directeur de thèse:	Shenjian Chen	Nanjing Univ.
Président:	Yuanning Gao	Tsinghua Univ.
Rapporteur:	Cristinel Diaconu	Aix Marseille Univ.
Rapporteur:	Haijun Yang	Shanghai Jiao Tong Univ.
Examinateur:	Yeuk-kwan Edna Cheung	Nanjing Univ.

Résumé

Le boson de Higgs est observé dans le mode de désintégration WW^* à travers un excès d'événements sur le bruit de fond de 6,1 écarts-types dans l'état final avec dilepton, alors que l'importance du signal attendu pour le boson de Higgs du modèle standard est de 5,8 écarts-types. Une indication pour la production du processus en fusion de bosons vecteurs (VBF) est également obtenue avec une importance de 3,2 écarts-types. Les résultats sont obtenus à partir d'un échantillon de données en collisions proton-proton enregistrées par le détecteur ATLAS au LHC, qui correspond à une luminosité intégrée de $4{,}5\,\text{fb}^{-1}$ à $\sqrt{s} = 7\,\text{TeV}$ et $20{,}3\,\text{fb}^{-1}$ à $\sqrt{s} = 8\,\text{TeV}$. Tous les canaux de saveur leptonique ($ee/\mu\mu$ et $e\mu/\mu e$) sont analysés, y compris de différentes catégories en multiplicité de jets ($n_j = 0, 1, \geq 2$). Les canaux ayant différentes saveurs leptoniques $e\mu/\mu e$ ont la meilleure sensibilité au signal. Les catégories à basse multiplicité de jets sont peuples principalement par la production du processus en fusion de gluon-gluon (ggF), tandis que la catégorie à haute multiplicité est plus sensible à la production VBF. Les bruits de fond dans différents canaux et catégories varient et leurs contributions sont obtenues dans la plupart de cas à partir des données avec des régions de contrôle ou validation. Les bruits de fond dominants sont les productions WW et le quark top. Dans les canaux ayant la même saveur leptonique, la contribution Drell-Yan est aussi une autre source importante. Pour le boson de Higgs à 125,36 GeV, le rapport du signal mesuré sur celui du modèle standard est de $1{,}09^{+0,16}_{-0,15}\,(\text{stat.})^{+0,17}_{-0,14}\,(\text{syst.})$. Les rapports correspondants pour les productions ggF et VBF sont de $1{,}02 \pm 0,19\,(\text{stat.})^{+0,22}_{-0,18}\,(\text{syst.})$ et $1{,}27^{+0,44}_{-0,40}\,(\text{stat.})^{+0,30}_{-0,21}\,(\text{syst.})$, respectivement. La section efficace totale mesurée à $\sqrt{s} = 8\,\text{TeV}$ est de $\sigma(gg \to H \to WW^*) = 4{,}6 \pm 0,9\,(\text{stat.})^{+0,8}_{-0,7}\,(\text{syst.})\,\text{pb}$ et $\sigma(\text{VBF}\,H \to WW^*) = 0{,}51^{+0,17}_{-0,15}\,(\text{stat.})^{+0,13}_{-0,08}\,(\text{syst.})\,\text{pb}$. La section efficace fiducielle est aussi mesurée pour la production ggF dans l'était final exclusif avec zéro ou un seul jet.

En plus des couplages, d'autres propriétés du boson de Higgs, notamment le nombre quantique de spin et la largeur totale de désintégration, sont également étudiées en utilisant les données de 8 TeV et les canaux ayant différentes saveurs leptoniques ($e\mu/\mu e$) seulement. L'étude du spin est basée sur un échantillon de données dominé par les événements sous le pic de résonance du boson de Higgs en utilisant les catégories de jets $n_j \leq 1$. L'hypothèse sur le spin-parité $J^{CP} = 0^{++}$ du modèle standard est comparée à d'autres hypothèses. Les données sont compatibles avec le modèle standard et les limites sont placées sur des hypothèses alternatives

de spin. Les événements dans la région à haute masse hors pic de résonance sans sélection sur la multiplicité de jets sont utilisés pour mesurer les couplages du boson de Higgs hors de sa couche de masse et pour imposer une contrainte sur la limite supérieure de la largeur totale du boson de Higgs, sous certaines hypothèses. Une amélioration de deux ordres de grandeur par rapport aux estimations directes est obtenue.

Mots-clés: Modèle Standard, Higgs, ggF, VBF, WW^*, spin, largeur.

Abstract

The Higgs boson decays to WW^* are observed based on an excess of events over background corresponding to 6.1 standard deviations in the dilepton final state including all lepton flavour channels ($ee/\mu\mu$ and $e\mu/\mu e$) and jet multiplicity categories ($n_j = 0, 1, \geq 2$), for a Standard Model expectation of 5.8 standard deviations. Evidence for the vector-boson fusion (VBF) production process is also obtained with a significance of 3.2 standard deviations. The different flavour channels $e\mu/\mu e$ have the best expected signal sensitivity. The low jet multiplicity categories are dominantly sensitive to the gluon-gluon fusion (ggF) production process whereas the large jet multiplicity category has the best sensitivity to the VBF production process. The results are obtained from proton-proton collision data recorded by the ATLAS detector at the LHC, corresponding to $4.5\,\text{fb}^{-1}$ at $\sqrt{s} = 7\,\text{TeV}$ and $20.3\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$. The background contribution in each channel and jet multiplicity category varies and is determined mostly with data-driven techniques with dedicated control or validation regions. The dominant background processes are the continuum WW and top quark productions. In the same flavour channels ($ee/\mu\mu$), the Drell-Yan process is another important background source. For a Higgs boson mass of 125.36 GeV, the ratio of the measured value to the expected value of the total production cross section times branching ratio fraction is $1.09^{+0.16}_{-0.15}\,(\text{stat.})^{+0.17}_{-0.14}\,(\text{syst.})$. The corresponding ratios for the gluon fusion and vector-boson fusion production modes are $1.02 \pm 0.19\,(\text{stat.})^{+0.22}_{-0.18}\,(\text{syst.})$ and $1.27^{+0.44}_{-0.40}\,(\text{stat.})^{+0.30}_{-0.21}\,(\text{syst.})$, respectively. At $\sqrt{s} = 8\,\text{TeV}$, the total production cross section is measured to be $\sigma(gg \to H \to WW^*) = 4.6 \pm 0.9\,(\text{stat.})^{+0.8}_{-0.7}\,(\text{syst.})\,\text{pb}$ and $\sigma(\text{VBF}\,H \to WW^*) = 0.51^{+0.17}_{-0.15}\,(\text{stat.})^{+0.13}_{-0.08}\,(\text{syst.})\,\text{pb}$. The fiducial cross section is determined for the gluon-fusion process in exclusive final states with zero and one associated jet.

In addition to the on-shell couplings, other properties of the Higgs boson, namely the spin quantum number and the total decay width, are also studied using the 8 TeV data and the different flavour channels only. The spin study is based on the on-shell dominated event sample using the $n_j \leq 1$ jet categories. The Standard Model spin-parity $J^{CP} = 0^{++}$ hypothesis is compared with alternative hypotheses. The data are found to be consistent with the Standard Model and limits are placed on alternative spin hypotheses. The off-shell events in the high mass tail from the inclusive jet category are then used to measure the off-shell couplings and impose a constraint on the upper limit of the total width of the Higgs boson

indirectly, when certain assumptions are made. The result improves by two orders of magnitude the direct estimations.

Keywords : Standard model, Higgs, ggF, VBF, WW^*, coupling, spin, width.

Contents

Chapter 1

Introduction

On 4 July 2012, a new particle with mass between 125 and 127 GeV, resembling the Higgs boson, was announced to be discovered by the ATLAS and CMS experiments [1, 2] at the Large Hadron Collider (LHC). The ATLAS discovery is best illustrated in Fig. 1.1 [1], which is the statistical analysis result of the ATLAS combined Higgs search consisting of contributions from three channels: WW^*, ZZ^*, and $\gamma\gamma$.

Since then, much efforts have been made to study the properties of this new particle and it is found that this particle behaves in many ways predicted by the Standard Model (SM) [3, 4]. However, more data and checks are still needed to know if it is the SM predicted Higgs or a boson in a larger framework Beyond Standard Model (BSM) Higgs.

In the thesis, the Higgs search in the WW^* channel at ATLAS is refined based on the original analysis contributing to the Higgs discovery Based on the refined analysis, the properties of the Higgs boson, like the on-shell/off-shell couplings and spin, are measured and found to be consistent with the Standard Model. It is an honour that I got the chance to be involved in all these analyses, which will be described in the following chapters.For those parts I have particularly contributed to, they will be mentioned in more detail.

1.1 Higgs boson

The Higgs boson, according to the SM, is the excitation of the Higgs field, which is introduced to explain the origin of mass, through the so-called Higgs mechanism to cause the spontaneously symmetry breaking (SSB) of electro-weak

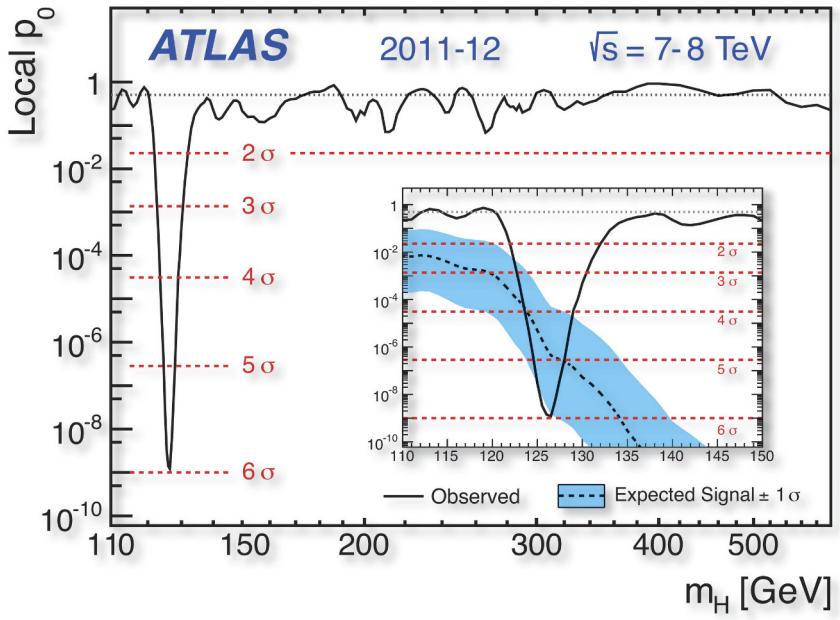


Figure 1.1: The result of the statistical analysis of the ATLAS combined Higgs search is shown in terms of local p_0 over Higgs mass. The local p_0 is to quantify the deviation from the non-Higgs Standard Model. A p_0 value close to unity means no significant deviation while a value being very small and away from unity indicates disagreement between data and model. The solid curve is for the observation while the dash for the expectation. Besides, the correspondences between the p_0 values and the number of standard deviations are marked out with red dash lines.

symmetry [5, 6]. Thus its existence is a crucial test to the SM and of whether our understanding of the SSB is correct.

Now we have found a neutral scalar particle.We are measuring its properties to test if it is the SM Higgs boson. Establishing the Higgs boson in the WW^* channel is also of great importance since it could give an answer to the question that whether or not fundamental scalar field exists in nature (in the framework of the SM, the answer is yes, but if we go to BSM, the Higgs could be a composition of other fundamental spin-1/2 particles). Moreover, which is beyond the topic of this thesis, the study of the Higgs boson could give us insight on the problems in cosmology, like inflation, the fate of the universe, the "cosmological constant" problem, etc.

1.2 $H \to WW^* \to \ell\nu\ell\nu$ channel

There are several Higgs decay modes that can be exploited to study the Higgs. In addition to the three channels mentioned above, there are other possibilities like $b\bar{b}$ and $\tau\tau$. Each decay mode has its own signature by which we can identify the Higgs boson.

As we will see, the WW^* channel dominants in a wide Higgs mass range, which makes it a perfect channel for the Higgs mass scanning. At the mass point of 125 GeV, its branching ratio is the second largest, with the $b\bar{b}$ channel being the first. However, requiring the two W bosons decaying leptonically makes it suffering much less backgrounds than the $b\bar{b}$ channel does.

The disadvantage of the WW^* channel is obvious: since the two neutrinos from the two leptonic decays of W bosons can not be detected by the detector, the Higgs mass cannot be reconstructed. This makes the channel less sensitive to the Higgs mass. For the same reason, the Higgs width information cannot be extracted directly from this channel, while a direct width measurement is possible in the ZZ^* and $\gamma\gamma$ channels [7]. But as we will see in later sections, an indirect width measurement is possible in the WW^* channel.

1.3 LHC and ATLAS

Until now, the LHC is the world's largest and most powerful particle collider. It is designed to collide two proton beams in its 27 kilometres tunnel at an unprecedented high centre of mass energy, and produce enormous amount of particles, in which new and exotic particles are being searched for.

Upon the LHC, there are four main detectors (and three small ones), which are the ATLAS, CMS, ALICE and LHCb. Two of them, the ATLAS and CMS, are general purpose particle detectors, while the rest two have more specific roles. It is on the ATLAS and CMS that a Higgs-like new particle was discovered.

ATLAS is a short name for "A Toroidal LHC Apparatus". And results presented in this thesis are all based on the data collected by this detector. There will be a more detailed description about the ATLAS detector as well as the LHC in later chapters.

1.4 Organisation of the thesis

The thesis is organised as follows:

Chapter 2: Introduction of the Standard Model, especially on the Higgs mechanism,

Chatper 3: Introduction of the LHC and the basic physics of proton-proton (pp) collision as well as description of the ATLAS component by component and the event reconstruction,

Chapter 4: Summary of the data and Monte Carlo samples used in the analyses to be presented in the thesis,

Chapter 5: Detail description of the observation of the Higgs boson in the $H \to WW^* \to \ell\nu\ell\nu$ channel as well as the on-shell Higgs coupling measurements,

Chapter 6: Description of the spin analysis in the $H \to WW^* \to \ell\nu\ell\nu$ channel,

Chapter 7: Description of the off-shell Higgs coupling measurement in the $H \to WW^* \to \ell\nu\ell\nu$ channel,

Chapter 8: The summary of the thesis,

Appendix A: Review of the data-driven methods of top background estimation used in the $H \to WW^* \to \ell\nu\ell\nu$ channel in ATLAS and CMS,

Appendix B: Introduction to the fake muon study in ATLAS.

1.5 Personally contributions

My personally contributions in the analyses to be presented in the thesis are summarised in the following list, together with the references to the corresponding chapters or sections (notations used here will be defined later in the thesis):

1. For the Higgs observation and (on-shell) coupling analysis, I have been one of the initiators for the inclusion of low-p_T events and have performed initial studies to show the significant gain (up to 20%) for the signal acceptance (Sec. 5.3.1). I have also been the responsible person for carrying out the top background estimation in the $n_j = 0$ channel (Sec. 5.4.2) and performing all relevant systematic uncertainty studies (Sec. 5.5.3). In particular the

theoretical systematic uncertainty used to be the dominant contribution in the top background estimation, I have made detailed studies and improved the uncertainty by a factor of two by generating a number of large statistics top MC samples at generator level and by using NLO MC generators instead of LO ones.

2. For the spin analysis, which is improved with respect to the previous publication by including the $n_j = 1$ channel to gain the sensitivity of discriminating the SM spin-0 hypothesis against alternative spin-2 hypotheses. I am the main contributor of this channel and has studied input variables for the BDT training (Sec 6.5) and systematic uncertainties of the top background in the channel (Sec. 6.6).

3. For the off-shell analysis, I am one of the fewer analysers (together with Jonathan Long and Yanping Huang). In particular I have validated the signal MC samples used for this analysis (Sec. 7.2) and performed the corresponding systematic uncertainty evaluation (Sec 7.5.4).

4. I am one of the authors of the top review paper and performed quantitative comparison of different methods in all the jet multiplicity channels (Appendix A). This study not only confirms that the current baseline choice for the coupling analysis in ATLAS is optimal but also provides guideline for future analyses.

5. Finally I have performed the fake muon study as my qualification task (Appendix B). This preliminary study of reducing fake muon rate, performed using a BDT technique with two different MC samples treating as signal muons reconstructed with the 3rd muon chain software to be used for RUN II, shows promising results. The same technique is now being applied in other RUN I analyses such as the $B_{(s)} \to \mu\mu$ rare decays.

Chapter 2

The Standard Model

In this chapter, we are going to describe the SM, which since about 1974 has successfully explained all the experimental observations at particle collider experiments. We will just cover the main ideas of the SM: the Gauge theory and the Higgs mechanism [8–11]. But there is tremendous detail we are skipping over (so topics like perturbative expansion, renormalisation, group representation etc. will not be discussed).

Besides, all the formulas and plots as well as the contents in this chapter are mainly based on two books: (a) "A Modern Introduction to Quantum Field Theory" by Michele Maggiore [5]; (b) "Symmetries and the Standard Model – Mathematics and Particle Physics" by Matthew Robinson [6].

2.1 Overview

The SM is a theory of the fundamental particles and fundamental forces. (excluding gravitation). All its contents are summarised in Fig. 2.1.

In terms of particles, there is only one scalar boson (spin-0 and neutral), the Higgs (in the yellow box), which gives masses to the rest massive particles in the figure. As for spin-1/2 fermions, there are two groups, which are leptons (in the green boxes) and quarks (in the purple boxes) depending on whether or not they take part in the strong interaction. Besides, each group consists of three generations: for leptons, each generation is composed of a lepton and its partner neutrino; for quarks, each generation is composed of a $+2/3$ charged quark and a $-1/3$ changed quark. As for spin-1 bosons, there are four types of them (in the red boxes and they are also called gauge bosons): which are the photon γ, the gluon

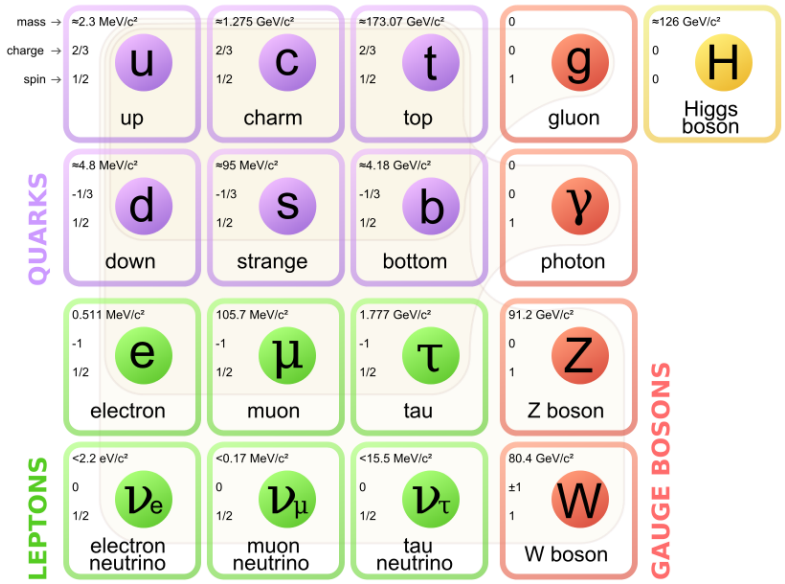


Figure 2.1: The particles in the SM.

g, the W boson and the Z boson.

In terms of interactions, there are three types and they are assumed to be propagated through exchange of the corresponding force-carrier particles, which are the spin-1 bosons we just mentioned. The electromagnetic force is propagated by the massless photon thus it is long-ranged. The weak force is propagated by the massive W and Z bosons so it is short-ranged. The strong force is propagated also by massless particles, the gluons. However, it is not long-ranged as in the case of EM force due to the fact that the gluons interact with themselves (while the photons don't).

If we look at the SM in a more mathematical way, its structure can be summarised in Table 2.1.

As the title of the card suggests, the SM is in fact a gauge theory: Certain gauge symmetries are assumed and described by the corresponding gauge groups $(SU(3)_C \times SU(2)_L \times U(1)_Y)$; Elementary particles move and interact respecting these symmetries and they are described by representations of the gauge group.

The C on $SU(3)_C$ stands for color, which means that the symmetry operation acts in the color space. The L on $SU(2)_L$ means it only acts on the left-handed states. (Left-handed states mean particles described by the left-handed Weyl spinor), The Y on $U(1)_Y$ stands for hypercharge and is to distinguish $U(1)_Y$ from the $U(1)$ for electromagnetism (EM).

Table 2.1: The structure of the SM.

A Yang-Mills (Gauge) Theory with Gauge Group $SU(3)_C \times SU(2)_L \times U(1)_Y$ with left-handed Weyl fields in three copies of the representation $(1, \mathbf{2}, -1/2) \oplus (1, 1, 1) \oplus (\mathbf{3}, \mathbf{2}, 1/6) \oplus (\bar{\mathbf{3}}, 1, -2/3) \oplus (\bar{\mathbf{3}}, 1, 1/3)$ where the first entry of each triplet is the $SU(3)_C$ representation, the next entry is the $SU(2)_L$ representation, and the last entry specifies the value of the $U(1)$ hypercharge Y, and a single copy of a complex scalar field in the representation $(1, \mathbf{2}, -1/2)$

2.2 Gauge theory

Since the SM is a gauge theory, we will explain what a gauge theory is in this section, taking the Quantum ElectroDynamics (QED) as an example.

Let's start with the Dirac spinor field Ψ, which describes the electron and positron. Its dynamics is governed by the Lagrangian, which takes the form:

$$\mathcal{L} = \bar{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi \,, \tag{2.1}$$

where $\bar{\Psi}$ is the Dirac adjoint of Ψ, defined by:

$$\bar{\Psi} = \Psi^\dagger \gamma^0 \,. \tag{2.2}$$

The γ matrices satisfy the *Clifford algebra*:

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \,, \tag{2.3}$$

and m is the mass of the electron.

There is a global $U(1)$ symmetry in the system (by global, it means that the symmetry transformation is universal everywhere in spacetime): If we impose the $U(1)$ transformation, which is actually multiplying a phase factor onto the fields:

$$\Psi \to e^{i\alpha}\Psi \quad \text{and} \quad \bar{\Psi} \to \bar{\Psi}\mathrm{e}^{-\mathrm{i}\alpha} \,, \tag{2.4}$$

where α is a constant. It is easy to show that the Lagrangian does not change:

$$\begin{aligned} \mathcal{L}' &= (\bar{\Psi}e^{-i\alpha})(i\gamma^\mu\partial_\mu - m)(e^{i\alpha}\Psi) && (2.5)\\ &= \bar{\Psi}(e^{-i\alpha}e^{i\alpha})(i\gamma^\mu\partial_\mu - m)\Psi && (2.6)\\ &= \bar{\Psi}(1)(i\gamma^\mu\partial_\mu - m)\Psi = \mathcal{L}\,. && (2.7)\end{aligned}$$

Now we want to improve the global symmetry to be local. First of all, we will replace the constant α with a function of space and time:

$$\alpha \to \alpha(x)\,, \tag{2.8}$$

where x represents the four dimensional spacetime coordinates.

But this makes the theory no longer symmetric under the local $U(1)$ transformation. More specifically, the ∂_μ will act on $\alpha(x)$ as well as on Ψ thus bring extra terms to the Lagrangian:

$$\begin{aligned} \mathcal{L}' &= (\bar{\Psi}e^{-i\alpha(x)})(i\gamma^\mu\partial_\mu - m)(e^{i\alpha(x)}\Psi) && (2.9)\\ &= \bar{\Psi}(e^{-i\alpha(x)}e^{i\alpha(x)})(i\gamma^\mu\partial_\mu - m - \gamma^\mu\partial_\mu\alpha(x))\Psi && (2.10)\\ &= \mathcal{L} - \bar{\Psi}\gamma^\mu\partial_\mu\alpha(x)\Psi\,. && (2.11)\end{aligned}$$

To recover the symmetry, we could process as follows: Introduce a vector field $A_\mu(x)$ which, under the local $U(1)$ transformation, transforms like:

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) - \frac{1}{e}\partial_\mu\alpha(x) \tag{2.12}$$

(the constant e is a convenient choice of normalisation, which will turn out to be the electric charge and the field A_μ will turn out to be the photon field). And then replace the ∂_μ in the expression of the Lagrangian with a so-called covariant derivative:

$$\partial_\mu \to D_\mu = \partial_\mu + ieA_\mu\,. \tag{2.13}$$

The formula here is actually a special case of the more general form:

$$D_\mu = \partial_\mu + igA^a_\mu T^a\,, \tag{2.14}$$

where g is the coupling constant of the interaction, T^a is the generator of the group and a is the generator indice. We can see that for each generator there will be a

corresponding gauge field. In the case of $U(1)$, we have only one generator which is the identity operator and thus only one vector field is introduced.

It is easy to check that the above covariant derivative, under the local $U(1)$ transformation, transforms in the same way as Ψ:

$$D_\mu \Psi = e^{i\alpha} D_\mu \Psi \,. \tag{2.15}$$

So the Lagrangian with this new derivative will become:

$$\begin{aligned} \mathcal{L}' &= (\bar{\Psi} e^{-i\alpha(x)})(i\gamma^\mu D_\mu - m)(e^{i\alpha(x)}\Psi) && (2.16)\\ &= \bar{\Psi}(e^{-i\alpha(x)} e^{i\alpha(x)})(i\gamma^\mu D_\mu - m)\Psi && (2.17)\\ &= \bar{\Psi}(1)(i\gamma^\mu \partial_\mu - m)\Psi = \mathcal{L}\,. && (2.18) \end{aligned}$$

Thus we recovered the local $U(1)$ symmetry!

If we expand the covariant derivative:

$$\begin{aligned} \mathcal{L}' &= \bar{\Psi}(i\gamma^\mu D_\mu - m)\Psi = \bar{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi - eA_\mu \bar{\Psi}\gamma^\mu \Psi && (2.19)\\ &= \mathcal{L}_{D,\text{free}} + \mathcal{L}_{\text{int}}\,, && (2.20) \end{aligned}$$

we get the Lagrangian for the free Dirac field and a new term which represents the coupling between the Dirac field and the introduced vector field. The constant e plays the role of coupling constant as expected.

To this step, we have revealed the essential part of the gauge theory: To promote the global symmetry (usually described by a Lie group and the group will be called **Gauge Group**) of the theory to be local, we could introduce additional fields (which will be called the **Gauge Fields**) with defined transformation property (which is designed to cancel out the terms that break the local symmetry) under the gauge group transformation, and replace the usual differential operators with the covariant derivative, in which we will put the gauge fields as well as introduce the coupling constants. This whole procedure is called **Gauge the Symmetry**. Expanding the covariant derivative, we get the terms in the Lagrangian that describe the interactions.

To be complete, we can add the kinematic term for the photon field and get the final Lagrangian for QED:

$$\mathcal{L} = \bar{\Psi}(i\gamma^\mu D_\mu - m)\Psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}\,, \tag{2.21}$$

where $F_{\mu\nu}$ is the electromagnetic tensor and is defined by:

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \,. \tag{2.22}$$

In summary, in a gauge theory, interactions are the results of gauging the symmetries, which are described by gauge groups, of the system.

2.3 Higgs mechanism

In this section, we will introduce another important component of the SM: the Higgs mechanism.

As it was shown in the previous section, we do not have a mass term for the photon field in the Lagrangian of QED. This is not due to the zero mass nature of the photon (rather, this is a consequence), but to the fact that a mass term for the gauge field will break the gauge symmetry of the theory (which can be easily checked).

It will not cause problem in the QED, since the photon has no mass. However, when we come to the weak interaction, the weak force carrier particles, like W^+, W^- and Z, have very large masses. In this case, we need to find other mechanism to acquire mass, which is the Higgs mechanism to be introduced in the following.

2.4 Spontaneous symmetry breaking

The Higgs mechanism is in fact a phenomenon of SSB, which is very general in nature. For example, the ferromagnetic materials, when they are below the Curie temperature, will develop a magnetisation in a particular direction spontaneously which breaks the space rotation symmetry. To understand exactly what the SSB is, we will use as an example the complex scalar field, whose Lagrangian is:

$$\mathcal{L} = -\frac{1}{2}\partial^\mu \phi^\dagger \partial_\mu \phi - V(\phi^\dagger, \phi)\,, \tag{2.23}$$

where the potential takes the form:

$$V(\phi^\dagger, \phi) = \frac{1}{2}\lambda m^2 (\phi^\dagger \phi - \Phi^2)^2\,. \tag{2.24}$$

Here λ and Φ are real constants.

It is easy to notice that the Lagrangian will have the global $U(1)$ symmetry. And if we display V versus ϕ, we get a "Mexican hat" potential shown in Fig. 2.2. The

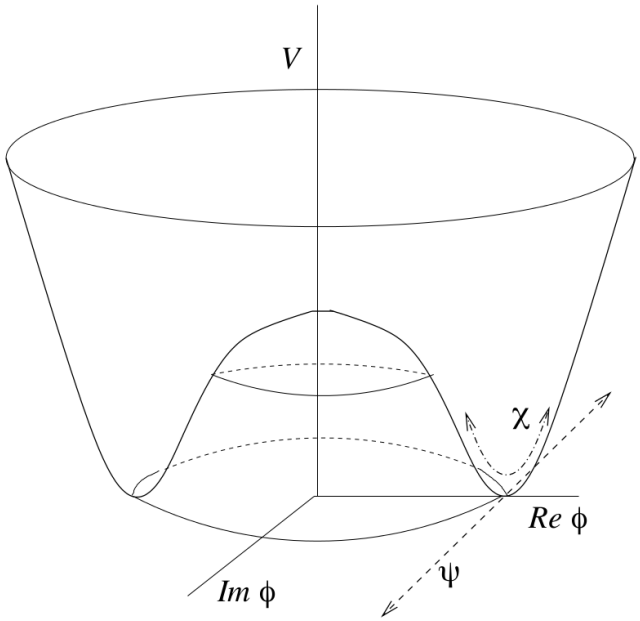


Figure 2.2: The "Mexican hat" potential of the complex scalar field ϕ and the two orthogonal fluctuations around a chosen vacuum.

vacuum, which is the field configuration that takes the minimum of the potential, is represented by the circle at $|\phi| = \Phi$. In other words, there are an infinite number of vacuums in this theory: one for every point on the circle at $|\phi| = \Phi$ and each of them are related by $U(1)$ transformation.

In order for the theory to make sense, we must choose a vacuum by hand. Since the choice is irrelevant due to the global $U(1)$ symmetry, we will choose the vacuum at $\phi = \Phi$, where ϕ is real. Now the system has no global $U(1)$ symmetry any more: by choosing a particular vacuum, we broke the symmetry.

Then we can expand the field around the chosen vacuum to study its perturbative behaviour: (the physical motivation for this expansion is that at low energies of a quantum field theory, particles are fluctuations around the vacuum)

$$\phi = \Phi + \chi + i\psi\,, \tag{2.25}$$

where χ and ψ are the new real scalar fields, which are shown in Fig. 2.2: one in the direction of the real axis and the other in the imaginary axis. Correspondingly,

the conjugation of the complex scalar field can be written as:

$$\phi^\dagger = \Phi + \chi - i\psi\,. \tag{2.26}$$

Now we can rewrite the Lagrangian with the new fields as:

$$\mathcal{L} = -\frac{1}{2}\partial^\mu(\chi - i\psi)\partial_\mu(\chi + i\psi) - \frac{1}{2}\lambda m^2[(\Phi + \chi - i\psi)(\Phi + \chi + 1\psi) - \Phi^2]^2 \tag{2.27}$$

$$= [-\frac{1}{2}\partial^\mu\chi\partial_\mu\chi - \frac{1}{2}4\lambda m^2\Phi^2\chi^2 - \frac{1}{2}\partial^\mu\psi\partial_\mu\psi] \tag{2.28}$$

$$-\frac{1}{2}\lambda m^2[4\Phi\chi^3 + 4\Phi\chi\psi^2 + \chi^4 + \chi^2\psi^2 + \psi^4]\,. \tag{2.29}$$

This is now a theory of a massive real scalar field χ (with mass $= \sqrt{4\lambda m^2\Phi^2}$), a massless real scalar field ψ, and five different types of interactions. The massive field χ is expected: since from Fig. 2.2, no matter it fluctuates left or right to the vacuum, it has to climb up a potential slope, which reflected in the Lagrangian as a mass term. In the same way, the massless ψ can be understood: since it moves in the groove with equal potential every where, it takes no effort at all to move freely, which is the feature of massless particle.

So we get the conclusion that breaking the symmetry (more precisely, global symmetry) results in the appearance of the massless field ψ. It turns out that breaking global symmetries as we have done always results in a massless boson, Such particles are called **Goldstone Bosons**.

2.5 Breaking local symmetries

In the previous section, we broke a global $U(1)$ symmetry. Now we will break a local U(1) and see what happens.

We begin with the Lagrangian for the complex scalar field again used in the previous section, but this time with the global $U(1)$ symmetry gauged to be local:

$$\mathcal{L} = -\frac{1}{2}(\partial^\mu - ieA^\mu)\phi^\dagger(\partial_\mu + ieA_\mu)\phi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - V(\phi^\dagger, \phi)\,, \tag{2.30}$$

where $V(\phi^\dagger, \phi)$ is again the Mexican hat potential.

Because the $U(1)$ symmetry is local now, we can choose the phase factor or the gauge $\alpha(x)$ so that not only is the vacuum real, but also ϕ is always real. Therefore

we get the expansion:

$$\phi = \Phi + h\,, \tag{2.31}$$

where h is a real scalar field representing fluctuations around the vacuum we chose in the direction of a real axis. Then the Lagrangian becomes:

$$\begin{aligned}
\mathcal{L} &= -\frac{1}{2}(\partial^\mu - ieA^\mu)(\Phi + h)^\dagger(\partial_\mu + ieA_\mu)(\Phi + h) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} && (2.32)\\
&\quad -\frac{1}{2}\lambda m^2[(\Phi + h)(\Phi + h) - \Phi^2]^2 && (2.33)\\
&= \ldots && (2.34)\\
&= -\frac{1}{2}\partial^\mu h\partial_\mu h - \frac{1}{2}4\lambda m^2\Phi^2 h^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}e^2\Phi^2 A^2 + \mathcal{L}_{\text{int}}\,, && (2.35)
\end{aligned}$$

where the allowed interaction terms include a vertex connecting an h and two A^μs, four hs, and three hs.

Before breaking the local symmetry, we had a complex scalar field ϕ and a massless vector field A^μ with two polarisation states. Now, we have a single real scalar h with mass $= \sqrt{4\lambda m^2\Phi^2}$ and a field A^μ with mass $= e\Phi$. In other words, our force-carrying particle A^μ has gained mass!

In summary, we started with a theory with no mass and by merely breaking the local symmetry, we have introduced mass into the theory. This mechanism for introducing mass into a theory is called the **Higgs Mechanism**. The resulting field h is called the **Higgs Boson**. Whereas the consequence of a global symmetry breaking is a massless boson called Goldstone boson, the consequence of a local symmetry breaking is that the gauge field, which came about as a result of the symmetry being local, acquires mass.

Moreover, we can generalise our conclusion to Non-abelian gauge groups with more than one generator: Gauge fields corresponding to broken generators acquire mass! The unbroken generators form a new gauge group that is smaller than the original group that was broken.

2.6 Gauge and Higgs sector

With the knowledge of the above two concepts (the gauge theory and the Higgs mechanism), we can start to construct the SM.

As the title of this section suggested, we will begin with the construction of the gauge and Higgs sector, which is the electroweak part of the SM gauge group

introduced in Fig. 2.1: the $SU(2)_L \times U(1)_Y$.

We start with the Higgs complex scalar doublet ϕ, which is the $(\mathbf{2}, -1/2)$ representation of $SU(2)_L \times U(1)_Y$ group (the $\mathbf{2}$ means it is a 2 dimensional representation of $SU(2)_L$ and $-1/2$ is the charge of the $U(1)_L$ part) and its Lagrangian is:

$$\mathcal{L} = -\frac{1}{2}\partial_\mu \phi_i^\dagger \partial^\mu \phi_i - V(\phi^\dagger, \phi)\,, \tag{2.36}$$

i stands for the doublet index. And $V(\phi^\dagger, \phi)$ is:

$$V(\phi^\dagger, \phi) = \frac{1}{4}\lambda(\phi^\dagger\phi - \frac{1}{2}\nu^2)^2\,, \tag{2.37}$$

which is a bit different from the one used in the examples of the previous section. But they are essentially the same except a change of the normalisation. And if we assume $\lambda > 0$, the minimum field configuration is at:

$$|\phi| = \frac{\nu}{\sqrt{2}}. \tag{2.38}$$

It is clear that the Lagrangian is symmetric under the global $SU(2)_L \times U(1)_Y$ transformation. And by gauging the symmetry, we could introduce gauge bosons and their interactions with the Higgs. To this end, we need to get the covariant derivative first using Formula 2.14.

Without further explanation, we write down the generators of the group:

$$T^a_{\mathbf{2}} = \frac{1}{2}\sigma^a\,, \tag{2.39}$$

where σ^a are the Pauli matrices. These are for the $SU(2)_L$ part and:

$$Y = C\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\,, \tag{2.40}$$

which is for the $U(1)_Y$ part (as mentioned before, Y is to distinguish from $U(1)$ that is gauged by the EM filed). C is the hypercharge and is $-1/2$ in this case (again, hypercharge is to distinguish from the electric charge).

If we denote the gauge fields to be introduced as W^a_μ corresponding to $T^a_{\mathbf{2}}$ and B_μ corresponding to Y, we get the covariant derivative:

$$(D_\mu\phi)_i = \partial_\mu\phi_i - i[g_2 W^a_\mu T^a_{\mathbf{2}} + g_1 B_\mu Y]_{ij}\phi_j\,, \tag{2.41}$$

where g_1 and g_2 are coupling constants for the $U(1)_Y$ and $SU(2)_L$, respectively.

Replacing the ∂_μ in the Lagrangian with D_μ, we have:

$$\mathcal{L} = -\frac{1}{2} D_\mu \phi_i^\dagger D^\mu \phi_i - V(\phi^\dagger, \phi)\,. \tag{2.42}$$

By expanding the covariant derivative, we can get the terms describing the interactions between the Higgs and the gauge bosons, which will not be given explicitly here.

Rather let's break the symmetry we just gauged (to gain mass for the gauge bosons). Any doublet satisfying Formula 2.38 can be chosen as the vacuum. But using a global $SU(2)_L$ transformation, we can always make the second component to be 0. And then make a global $U(1)_Y$ transformation to make the field real. Therefore, the vacuum can be written as:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} \nu \\ 0 \end{pmatrix}\,. \tag{2.43}$$

We used subscript 0 to indicate this is for the vacuum configuration of the field where there is nothing. So we can expand ϕ around this new vacuum:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \nu + h(x) \\ 0 \end{pmatrix}\,. \tag{2.44}$$

There is only one real scalar field $h(x)$ in the expansion, because we have chosen our $SU(2)_L$ gauge to keep the second component 0 and our $U(1)_Y$ gauge to keep the first component real (this gauge choice is called the unitarity gauge).

Plugging this into the Lagrangian and collecting together the terms related to ν, we have:

$$\mathcal{L}_0 = -\frac{1}{8}\nu^2 V_\mu^T \begin{pmatrix} g_2^2 & 0 & 0 & 0 \\ 0 & g_2^2 & 0 & 0 \\ 0 & 0 & g_2^2 & -g_1 g_2 \\ 0 & 0 & -g_1 g_2 & g_1^2 \end{pmatrix} V^\mu\,, \tag{2.45}$$

where $V_\mu^T = (W_\mu^1, W_\mu^2, W_\mu^3, B_\mu)$. We only write out the terms related to ν because all the mass information of the gauge bosons is included in the terms that contain it.

Then we do the following redefinition of the gauge fields:

$$W_\mu^+ = \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2) \tag{2.46}$$

$$W_\mu^- = \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2) \tag{2.47}$$

$$Z_\mu = \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu \tag{2.48}$$

$$A_\mu = \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu\,, \tag{2.49}$$

where θ_W is called the weak mixing angle defined by:

$$\theta_W = \tan^{-1}\left(\frac{g_1}{g_2}\right)\,. \tag{2.50}$$

This redefinition is nothing but a rotation of the basis. But it has two advantages: First, the mass matrix is now fully diagonalised. Especially, the mass corresponds to the new gauge field A_μ is 0. Second, the final Lagrangian will have an explicit $U(1)$ symmetry. All these indicate that we have broken the original $SU(2)_L \times U(1)_Y$ symmetry to a smaller $U(1)$ symmetry, which corresponds to the fact that the photon is still massless. Thus we denote this $U(1)$ as $U(1)_{\rm EM}$.

After the above rotation, Formula 2.45 becomes:

$$\mathcal{L}_0 = -M_W^2 W^{+\mu} W_\mu^- - \frac{1}{2} M_Z^2 Z^\mu Z_\mu\,, \tag{2.51}$$

where we have defined:

$$M_W = \frac{g_2 \nu}{2}, \quad \text{and} \quad M_Z = \frac{M_W}{2\cos\theta_W} = \frac{\nu}{2}\sqrt{g_1^2 + g_2^2}\,. \tag{2.52}$$

Thus, through symmetry breaking, we have gained mass for W_μ^+, W_μ^-, and Z_μ bosons,

To this step (we will leave out the work of writing down kinematic terms for the gauge fields for simplicity) , we have constructed the theory describing the Higgs, the weak force carriers (Ws and Z) and the EM force carrier (the photon), as well as the interactions among them.

In summary, at very high energies (above the breaking of the $SU(2)_L \times U(1)_Y$), we have a Higgs complex scalar field and 4 massless vector boson fields $(W_\mu^1, W_\mu^2, W_\mu^3, B_\mu)$. At low energies, however, the symmetry is broken, and the low energy effective theory consists of a linear combination of the original four gauge fields. Three of them

have gained mass and one remains massless. The theory above the symmetry breaking scale is called **Electroweak Theory**. Whereas below the breaking scale they become two separate forces: the broken weak and unbroken electromagnetic. This is the first and most basic example of the unification theory: at low energies, the electromagnetic and weak forces are separate. At high energies, they unify into a single theory that is described by $SU(2)_L \times U(1)_Y$.

2.7 Lepton sector

Now we turn to the lepton sector, which is still in the $SU(2)_L \times U(1)_Y$ part of the SM gauge group. According to the previous description, there are six **Flavours** of leptons arranged into three **Families** or **Generations**. Each family behaves nearly exactly the same way, so we only discuss one generation, and then make three copies of the structure for the three generations, and allow mixing between them.

First of all, there is an asymmetry to be noticed: only the left-handed fields are included in the weak interactions and there exists only left-handed neutrino, while for the electron, there exists both left-handed and right-handed states.

Thus we start with the following representation: the left-handed neutrino is grouped together with the left-handed electron into a $SU(2)_L$ doublet:

$$L = \begin{pmatrix} \nu_e \\ e \end{pmatrix} . \tag{2.53}$$

And notice that a right-handed electron is the same as the left-handed positron, so we include into our theory another left-handed $SU(2)_L$ singlet, denoted as $\bar{e}$. In short, we have two representations, one $(\mathbf{2}, -1/2)$ and the other $(\mathbf{1},1)$.

Mimicking what we did in the previous section, we can write down the covariant derivative for both representations:

$$(D_\mu L)_i = \partial_\mu L_i - i g_2 W^a_\mu (T^a)_{ij} L_j - i g_1 B_\mu Y_L L_i \tag{2.54}$$

$$D_\mu \bar{e} = \partial_\mu \bar{e} - i g_1 B_\mu Y_{\bar{e}} \bar{e} \,, \tag{2.55}$$

where Y_L and $Y_{\bar{e}}$ are:

$$Y_L = -\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad Y_{\bar{e}} = (1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{2.56}$$

The field $\bar{e}$ has no $SU(2)$ term in its covariant derivative because the **1** representation of $SU(2)$ is the trivial representation.

By replacing the ∂_μ with the covariant derivative in the kinematic terms of the Lagrangian for the spin-1/2 fields, we have:

$$\mathcal{L}_{kin} = iL^{\dagger i}\bar{\sigma}^\mu(D_\mu L)_i + i\bar{e}^\dagger\bar{\sigma}^\mu D_\mu\bar{e}\,, \tag{2.57}$$

and expanding the covariant derivative, we can get the interactions between the fermions and gauge bosons as usual.

We will look at some example interaction terms, rather than list all of them. After we do the same rotation as in the previous section, we have:

$$g_2(W_\mu^1 T^1 + W_\mu^2 T^2) = \frac{g_2}{\sqrt{2}}\begin{pmatrix} 0 & W_\mu^+ \\ W_\mu^- & 0 \end{pmatrix}. \tag{2.58}$$

This gives the e, ν_e and W vertex in the Feynman diagrams. More mathematically, if we treat the $(\nu_e, 0)^T$ and $(0, e)^T$ as the basis of the **2** representation of the $SU(2)$ group, with isospin charge $1/2$ and $-1/2$ respectively (isospin could be thought of as the $SU(2)$ charge, similar as the electric charge of $U(1)_{\text{EM}}$). The W^+ boson will raise the isospin charge from $-1/2$ to $1/2$ (while W^- will lower the charge by a unit). Thus the interaction is called charged current interaction.

Another example is:

$$g_2 W_\mu^3 T^3 + g_1 B_\mu Y = e(T^3 + Y)A_\mu + e(\cot\theta_W T^3 - \tan\theta_W Y)Z_\mu\,. \tag{2.59}$$

The term containing A_μ gives the EM interaction vertex, with $T^3 + Y$ quantifying the charge of the interacting particle. Thus we define the generator of the electric charge to be $Q = T^3 + Y$, and since we know exactly the matrix of Q, it is easy to check that we have:

$$Q\nu_e = 0, \quad Qe = -e, \quad Q\bar{e} = +\bar{e}\,, \tag{2.60}$$

as expected. The term containing Z_μ, not like the case of W^+ or W^-, will not change the isospin charge, thus the corresponding interaction is called neutral current interaction.

Now we try to add a mass term for our lepton fields. But it is observed that we cannot have a mass term satisfying both Lorentz invariance and gauge invariance. For example, the term $\bar{e}^a\bar{e}^b\epsilon_{ab} = (\bar{e}\bar{e})$ (ϵ^{ij} is the totally antisymmetric tensor) is both Lorentz invariant and $SU(2)_L$ invariant, but not $U(1)_Y$ invariant, since $\bar{e}$ has

a hypercharge of +1 thus two $\bar{e}$s have a hypercharge of +2.

So we turn to the Higgs mechanism again for help. We can add a Yukawa term:

$$\mathcal{L}_{Yuk} = -y\epsilon^{ij}\phi_i(L_j\bar{e}) + h.c.\,, \tag{2.61}$$

where y is the Yukawa coupling constant, and "*h.c.*" is the Hermitian conjugate of the first term. This term has a hypercharge of zero and the $SU(2)_L$ and Lorentz indices are all contracted to form singlets.

Then we just repeat what we did, breaking the symmetry as in the previous section: choose the unitarity gauge and replace ϕ with its fluctuation around its vacuum using Formula 2.44, so we get (eliminating the details):

$$\mathcal{L}_{Yuk} = -\frac{1}{\sqrt{2}}y\nu\bar{\varepsilon}\varepsilon - \frac{1}{\sqrt{2}}y\nu h\bar{\varepsilon}\varepsilon,\,, \tag{2.62}$$

where

$$\varepsilon = \begin{pmatrix} e \\ i\sigma^2\bar{e}^\dagger \end{pmatrix}, \tag{2.63}$$

is the Dirac field for the electron. Now we get a mass term for the electron and positron where $m_e = y\nu/\sqrt{2}$. Besides, the field e and $i\sigma^2\bar{e}^\dagger$ which are initially unrelated now join together to form the left- and right-handed parts of the electron field ε.

To summarise, we used a trick to incorporate the asymmetry between left- and right-handed spinors when they take part in the weak interaction. We start from the free theory for the spin-1/2 fields and by gauging the theory, we introduced the interactions between the gauge bosons and the fermions. Finally, we used the Higgs mechanism to once again give fermions masses.

2.8 Quark sector

A quark is a spin-1/2 particle that interacts with the $SU(3)$ color force. Just as leptons, there are 6 flavours of quarks, arranged into three families or generations. As before, we will work with only one generation, which are the *up* and *down* quarks.

We start with three fields: Q, $\bar{u}$ and d, which are the ($\mathbf{3}$,$\mathbf{2}$,+1/6),($\bar{\mathbf{3}}$,1,−2/3), and ($\bar{\mathbf{3}}$,1,+1/2) representations of $SU(3)_C \times SU(2)_L \times SU(1)_Y$ group ($\bar{\mathbf{3}}$ is the conjugate representation of $\mathbf{3}$ with opposite charges to it). We used the same trick

here: the left-handed parts of the two quarks in the same family are put together into a $SU(2)$ doublet:

$$Q = \begin{pmatrix} u \\ d \end{pmatrix}, \tag{2.64}$$

while their right-handed parts are represented by the left-handed parts of their anti-particles $\bar{u}$ and $\bar{d}$.

Again we can write out the covariant derivative for all three fields:

$$\begin{aligned}
(D_\mu Q)_{\alpha i} &= \partial_\mu Q_{\alpha i} - ig_3 A^a_\mu (T^a_3)^\beta_\alpha - ig_2 W^a_\mu (T^a_2)^j_i Q_{\alpha j} - ig_1 (\frac{1}{6}) B_\mu Q_{\alpha i} && (2.65)\\
(D_\mu \bar{u})^\alpha &= \partial_\mu \bar{u}^\alpha - ig_3 A^a_\mu (T^a_{\bar{3}})^\alpha_\beta \bar{u}^\beta - ig_1 (-\frac{2}{3}) B_\mu \bar{u}^\alpha && (2.66)\\
(D_\mu \bar{d})^\alpha &= \partial_\mu \bar{d}^\alpha - ig_3 A^a_\mu (T^a_{\bar{3}})^\alpha_\beta \bar{d}^\beta - ig_1 (-\frac{2}{3}) B_\mu \bar{d}^\alpha , && (2.67)
\end{aligned}$$

where i is the $SU(2)_L$ indices and α, β the $SU(3)_C$ indices. The vector field A^a_μ is the gluon field, to be distinguished from the photon field.

Writing out the $SU(2) \times U(1)$ part of the covariant derivatives in terms of the W^+_μ, W^-_μ, Z_μ and A_μ introduced previously, we get the interactions between these gauge bosons and the quarks (for example, the u, d and W vertex). The procedure is almost the same so we do not go through it again, but just mention that using the generator of the electric charge Q we defined, we find as expected:

$$Qu = +\frac{2}{3}u, \quad Qd = -\frac{1}{3}d, \quad Q\bar{u} = -\frac{2}{3}\bar{u}, \quad Q\bar{d} = +\frac{1}{3}\bar{d}. \tag{2.68}$$

Let's now look at the $SU(3)$ part. Since there are eight generators for the $SU(3)$ group, we need to introduce eight gauge fields, which will be called gluons, to gauge the symmetry. And if we expand the covariant derivative, we will get the terms that describe the interactions between gluons and quarks.

For simplicity, we will not list every possibility, but just show an example: There is one generator T^a having the following form:

$$T^a = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.69}$$

If it acts on, in the $SU(3)$ space, the vector $(0,1,0)^T$, it will give the vector $(1,0,0)^T$. And if we denote the three vectors $(1,0,0)^T$, $(0,1,0)^T$ and $(0,0,1)^T$ to be *red*, *green* and *blue* (r, g, b), this means that the gluon will destroy a *green* quark and produce

a *red* quark. Thus we can denote the gluon to be $g_r^{\bar{g}}$. There will be six such kinds of gluons which destroy one color of quark and produce another quark of different color (since there are in total three colors). These gluons can be compared with the W bosons in the previous section that change the weak isospin charge, except here what the gluons altered is the color charge of the strong interaction. As for the rest two gluons, they will not change the color of the state. Similarly, these are to be compared with the Z boson.

As for the masses of the quarks, just like in the case of leptons, we cannot write down a gauge invariant mass term but include a Yukawa term coupling these fields to the Higgs:

$$\mathcal{L}_{Yuk} = -y'\epsilon^{ij}\phi_i Q_{\alpha j}\bar{d}^{\alpha} - y''\phi^{\dagger i}Q_{\alpha i}\bar{u}^{\alpha} + h.c.\,, \tag{2.70}$$

where y' and y'' are the coupling constants. After we break the symmetry according to Formula 2.44, we get (we leave out the details):

$$\mathcal{L}_{Yuk} = -\frac{1}{\sqrt{2}}y'(\nu+h)\bar{D}^{\alpha}D_{\alpha} - \frac{1}{\sqrt{2}}y''(\nu+h)\bar{U}^{\alpha}U_{\alpha}\,, \tag{2.71}$$

where α is the color indice. D_α and U_α are the Dirac fields for the *up* and *down* quarks defined by:

$$D_\alpha = \begin{pmatrix} d_\alpha \\ \bar{i}\sigma^2 d_\alpha^* \end{pmatrix}, \quad U_\alpha = \begin{pmatrix} u_\alpha \\ \bar{i}\sigma^2 u_\alpha^* \end{pmatrix}. \tag{2.72}$$

And the masses for the *up* and *down* quarks are:

$$m_d = \frac{y'\nu}{\sqrt{2}}, \quad m_u = \frac{y''\nu}{\sqrt{2}}\,. \tag{2.73}$$

Again, the spontaneous symmetry breaking of ϕ_i links a term in the doublet with the singlet $\bar{u}$ and $\bar{d}$.

In conclusion, in the construction of the quark sector: at first, we put the left-handed parts of the two quarks in the same family together, and the right-handed parts separately. By gauging the $SU(3)$ symmetry, we introduced gluons as well as their interactions with quarks. Then by breaking the symmetry, we not only got the mass terms for the quarks but also linked the left- and right-handed parts of the same quark.

2.9 Yukawa couplings among generations

Since there are three generations of leptons and quarks, we will make three copies of the above lepton and quark structures: one for each generation. It is natural to generalise the theory to contain Yukawa couplings between generations. If we denote the generations with indices A,B,C, etc. The Yukawa coupling now becomes:

$$\mathcal{L}_{\text{Yuk}} = -Y^d_{AB}\phi Q^A \bar{d}^B - Y^u_{AB}\phi^\dagger Q^A \bar{u}^B - Y^e_{AB}\phi L^A \bar{e}^B + h.c.\,, \tag{2.74}$$

where we have suppressed the $SU(3)_C$, $SU(2)_L$ and the Lorentz indices. The u and d are general notations for the two quarks in a certain generation. And the e should be understood as lepton, rather than electron, of a lepton generation. Besides, it is easy to generalise the above couplings to contain the non-zero mass effect of neutrinos. But we stick to the original approximation where the neutrinos are massless.

The kinetic terms for the three generations of leptons $L^{A\dagger} i\sigma^\mu D_\mu L^A$ are invariant under global $SU(3)$ rotations of the fields $L'^A = R^A_B L^B$. The same is true for the three generations of quarks. Using this freedom, we can always rotate L and $\bar{e}$ to make the coupling Y^e_{AB} diagonalised with respect to indices A and B, which means that we can always eliminate the mixing of generations of leptons by rotation or redefinition. However, if we diagonalise Y^u_{AB} by rotating or redefining Q and $\bar{u}$, we are left without enough freedom to diagonalise Y^d_{AB}. Thus the mixing among generations cannot be cleaned and this observation predicts that the quarks will decay from heavier to lighter ones directly. This mixing is represented by the so-called CKM matrix.

By now, we have presented all the content of the SM. In the following chapters, if not mentioned specifically, we will limit ourself in the content of the SM.

Before moving on to the physics beyond the SM, we would like to reveal the whole picture again in terms of the full Lagrangian in Fig. 2.3.

2.10 Beyond the Standard Model

The SM is a rather unsatisfying end to a quest for understanding the fundamental nature of "everything". First of all, it does not describe everything like the masses of the neutrinos. Second, there are 19 free parameters (or even more if we

$$
\begin{aligned}
&\mathcal{L}_{\text{StandardModel}} = \\
&-\tfrac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \tfrac{1}{4}g_s^2 f^{abc} f^{ade} g^b_\mu g^c_\nu g^d_\mu g^e_\nu + \\
&\tfrac{1}{2}ig_s^2(\bar{q}^\sigma_i \gamma^\mu q^\sigma_j) g^a_\mu + \bar{G}^a \partial^2 G^a + g_s f^{abc}\partial_\mu \bar{G}^a G^b g^c_\mu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu - \\
&M^2 W^+_\mu W^-_\mu - \tfrac{1}{2}\partial_\nu Z^0_\mu \partial_\nu Z^0_\mu - \tfrac{1}{2c_w^2} M^2 Z^0_\mu Z^0_\mu - \tfrac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - \tfrac{1}{2}\partial_\mu H \partial_\mu H - \\
&\tfrac{1}{2}m_h^2 H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - M^2 \phi^+ \phi^- - \tfrac{1}{2}\partial_\mu \phi^0 \partial_\mu \phi^0 - \tfrac{1}{2c_w^2} M \phi^0 \phi^0 - \beta_h [\tfrac{2M^2}{g^2} + \\
&\tfrac{2M}{g} H + \tfrac{1}{2}(H^2 + \phi^0 \phi^0 + 2\phi^+ \phi^-)] + \tfrac{2M^4}{g^2}\alpha_h - igc_w[\partial_\nu Z^0_\mu (W^+_\mu W^-_\nu - \\
&W^+_\nu W^-_\mu) - Z^0_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + Z^0_\mu (W^+_\nu \partial_\nu W^-_\mu - \\
&W^-_\nu \partial_\nu W^+_\mu)] - igs_w[\partial_\nu A_\mu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - A_\nu (W^+_\mu \partial_\nu W^-_\mu - \\
&W^-_\mu \partial_\nu W^+_\mu) + A_\mu (W^+_\nu \partial_\nu W^-_\mu - W^-_\nu \partial_\nu W^+_\mu)] - \tfrac{1}{2}g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \\
&\tfrac{1}{2}g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2 c_w^2 (Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu - Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + \\
&g^2 s_w^2 (A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + g^2 s_w c_w [A_\mu Z^0_\nu (W^+_\mu W^-_\nu - \\
&W^+_\nu W^-_\mu) - 2A_\mu Z^0_\mu W^+_\nu W^-_\nu] - g\alpha[H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-] - \\
&\tfrac{1}{8}g^2\alpha_h[H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2 H^2] - \\
&gMW^+_\mu W^-_\mu H - \tfrac{1}{2}g\tfrac{M}{c_w^2} Z^0_\mu Z^0_\mu H - \tfrac{1}{2}ig[W^+_\mu(\phi^0\partial_\mu\phi^- - \phi^-\partial_\mu\phi^0) - \\
&W^-_\mu(\phi^0\partial_\mu\phi^+ - \phi^+\partial_\mu\phi^0)] + \tfrac{1}{2}g[W^+_\mu(H\partial_\mu\phi^- - \phi^-\partial_\mu H) - W^-_\mu(H\partial_\mu\phi^+ - \\
&\phi^+\partial_\mu H)] + \tfrac{1}{2}g\tfrac{1}{c_w}(Z^0_\mu(H\partial_\mu\phi^0 - \phi^0\partial_\mu H) - ig\tfrac{s_w^2}{c_w} M Z^0_\mu (W^+_\mu\phi^- - W^-_\mu\phi^+) + \\
&igs_w M A_\mu (W^+_\mu\phi^- - W^-_\mu\phi^+) - ig\tfrac{1-2c_w^2}{2c_w} Z^0_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) + \\
&igs_w A_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) - \tfrac{1}{4}g^2 W^+_\mu W^-_\mu [H^2 + (\phi^0)^2 + 2\phi^+\phi^-] - \\
&\tfrac{1}{4}g^2\tfrac{1}{c_w^2} Z^0_\mu Z^0_\mu [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2\phi^+\phi^-] - \tfrac{1}{2}g^2\tfrac{s_w^2}{c_w} Z^0_\mu \phi^0 (W^+_\mu\phi^- + \\
&W^-_\mu\phi^+) - \tfrac{1}{2}ig^2\tfrac{s_w^2}{c_w} Z^0_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) + \tfrac{1}{2}g^2 s_w A_\mu \phi^0 (W^+_\mu\phi^- + \\
&W^-_\mu\phi^+) + \tfrac{1}{2}ig^2 s_w A_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) - g^2\tfrac{s_w}{c_w}(2c_w^2 - 1) Z^0_\mu A_\mu \phi^+\phi^- - \\
&g^1 s_w^2 A_\mu A_\mu \phi^+\phi^- - \bar{e}^\lambda(\gamma\partial + m^\lambda_e)e^\lambda - \bar{\nu}^\lambda\gamma\partial\nu^\lambda - \bar{u}^\lambda_j(\gamma\partial + m^\lambda_u)u^\lambda_j - \\
&\bar{d}^\lambda_j(\gamma\partial + m^\lambda_d)d^\lambda_j + igs_w A_\mu[-(\bar{e}^\lambda\gamma^\mu e^\lambda) + \tfrac{2}{3}(\bar{u}^\lambda_j\gamma^\mu u^\lambda_j) - \tfrac{1}{3}(\bar{d}^\lambda_j\gamma^\mu d^\lambda_j)] + \\
&\tfrac{ig}{4c_w} Z^0_\mu[(\bar{\nu}^\lambda\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar{e}^\lambda\gamma^\mu(4s_w^2 - 1 - \gamma^5)e^\lambda) + (\bar{u}^\lambda_j\gamma^\mu(\tfrac{4}{3}s_w^2 - \\
&1 - \gamma^5)u^\lambda_j) + (\bar{d}^\lambda_j\gamma^\mu(1 - \tfrac{8}{3}s_w^2 - \gamma^5)d^\lambda_j)] + \tfrac{ig}{2\sqrt{2}} W^+_\mu[(\bar{\nu}^\lambda\gamma^\mu(1+\gamma^5)e^\lambda) + \\
&(\bar{u}^\lambda_j\gamma^\mu(1+\gamma^5)C_{\lambda\kappa}d^\kappa_j)] + \tfrac{ig}{2\sqrt{2}} W^-_\mu[(\bar{e}^\lambda\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar{d}^\kappa_j C^\dagger_{\lambda\kappa}\gamma^\mu(1 + \\
&\gamma^5)u^\lambda_j)] + \tfrac{ig}{2\sqrt{2}}\tfrac{m^\lambda_e}{M}[-\phi^+(\bar{\nu}^\lambda(1-\gamma^5)e^\lambda) + \phi^-(\bar{e}^\lambda(1+\gamma^5)\nu^\lambda)] - \\
&\tfrac{g}{2}\tfrac{m^\lambda_e}{M}[H(\bar{e}^\lambda e^\lambda) + i\phi^0(\bar{e}^\lambda\gamma^5 e^\lambda)] + \tfrac{ig}{2M\sqrt{2}}\phi^+[-m^\kappa_d(\bar{u}^\lambda_j C_{\lambda\kappa}(1-\gamma^5)d^\kappa_j) + \\
&m^\lambda_u(\bar{u}^\lambda_j C_{\lambda\kappa}(1+\gamma^5)d^\kappa_j] + \tfrac{ig}{2M\sqrt{2}}\phi^-[m^\lambda_d(\bar{d}^\lambda_j C^\dagger_{\lambda\kappa}(1+\gamma^5)u^\kappa_j) - m^\kappa_u(\bar{d}^\lambda_j C^\dagger_{\lambda\kappa}(1- \\
&\gamma^5)u^\kappa_j] - \tfrac{g}{2}\tfrac{m^\lambda_u}{M}H(\bar{u}^\lambda_j u^\lambda_j) - \tfrac{g}{2}\tfrac{m^\lambda_d}{M}H(\bar{d}^\lambda_j d^\lambda_j) + \tfrac{ig}{2}\tfrac{m^\lambda_u}{M}\phi^0(\bar{u}^\lambda_j\gamma^5 u^\lambda_j) - \\
&\tfrac{ig}{2}\tfrac{m^\lambda_d}{M}\phi^0(\bar{d}^\lambda_j\gamma^5 d^\lambda_j) + \bar{X}^+(\partial^2 - M^2)X^+ + \bar{X}^-(\partial^2 - M^2)X^- + \bar{X}^0(\partial^2 - \\
&\tfrac{M^2}{c_w^2})X^0 + \bar{Y}\partial^2 Y + igc_w W^+_\mu(\partial_\mu\bar{X}^0 X^- - \partial_\mu\bar{X}^+ X^0) + igs_w W^+_\mu(\partial_\mu\bar{Y}X^- - \\
&\partial_\mu\bar{X}^+ Y) + igc_w W^-_\mu(\partial_\mu\bar{X}^- X^0 - \partial_\mu\bar{X}^0 X^+) + igs_w W^-_\mu(\partial_\mu\bar{X}^- Y - \\
&\partial_\mu\bar{Y}X^+) + igc_w Z^0_\mu(\partial_\mu\bar{X}^+ X^+ - \partial_\mu\bar{X}^- X^-) + igs_w A_\mu(\partial_\mu\bar{X}^+ X^+ - \\
&\partial_\mu\bar{X}^- X^-) - \tfrac{1}{2}gM[\bar{X}^+X^+H + \bar{X}^-X^-H + \tfrac{1}{c_w^2}\bar{X}^0X^0H] + \\
&\tfrac{1-2c_w^2}{2c_w} igM[\bar{X}^+X^0\phi^+ - \bar{X}^-X^0\phi^-] + \tfrac{1}{2c_w} igM[\bar{X}^0X^-\phi^+ - \bar{X}^0X^+\phi^-] + \\
&igMs_w[\bar{X}^0X^-\phi^+ - \bar{X}^0X^+\phi^-] + \tfrac{1}{2}igM[\bar{X}^+X^+\phi^0 - \bar{X}^-X^-\phi^0].
\end{aligned}
$$

Figure 2.3: The full Lagrangian of the SM.

account for neutrino masses) in the model. To the reductionist physicist, so many free parameters suggest a more fundamental theory may exist. Thus physics BSM searches for explaining things that the SM cannot answer and for reduction in free parameters. And we will introduce the various BSMs in the following briefly.

BSM has taken many forms over the years. The main tracks currently popular are massive neutrinos, technicolor, supersymmetry, extra-dimensional models, grand unified theories, string theory, loop quantum gravity etc.

The observed neutrino mass has led to many models linking massive neutrinos to the SM: some where neutrinos have a Dirac mass, others Majorana mass, and others with a mix. The more popular models link the neutrino's small mass to the large mass scales of grand unified theories through a so-called "see-saw" mechanism.

Because fundamental scalars, like the Higgs in the SM, have poor renormalisation properties, technicolor and supersymmetry were suggested. Technicolor models give the W and Z mass from a 5th force (the technicolor force) and do not require a fundamental Higgs. Rather the Higgs is produced as a composite scalar formed from a condensate of spin-1/2 fermions that carry charges of an additional $SU(n)$, with $n > 3$, gauge symmetry.

A second alternative to a fundamental Higgs came from the string theory and then from brane models. Models with extra dimensions might not (and need not) carry an actual Higgs scalar. Instead, the role could be played by the 5th component of a gauge field (that acts like a scalar in our 4 dimensional spacetime).

Supersymmetry introduces a set of superpartner particles to compensate for the poor renormalisation properties of the Higgs. For every standard scalar, a superpartner spin-1/2 fermion is introduced. For every spin-1/2 matter fermion, a superpartner scalar is introduced. Furthermore, for every type of spin-1 gauge boson, there is a corresponding spin-1/2 "gaugino", while for the spin-2 graviton, there is an associated spin-3/2 "gravitino".

Large extra dimensional models either add dimensions to spacetime that are coiled up or confine matter to a 4-D slice of the larger dimensional space.

Grand unified theories (GUTs) are efforts to combine the three gauge coupling parameters into one.

Finally, the loop quantum gravity and string theory each try to combine quantum physics with gravity and provide an explanation for some or all of the above puzzles.

Chapter 3

Large Hadron Collider and ATLAS detector

The LHC is a huge proton-proton circular collider located at CERN, which opens a new frontier in particle physics due to its higher collision energy and luminosity compared to all the existing accelerators. And the ATLAS is one of the four main detectors built upon it, which is a general purpose detector with the guiding principle of maximising the discovery potential for new physics (such as Higgs bosons and supersymmetric particles) and keeping the capability of high-accuracy measurements of known objects (such as leptons, photons, quark and gluon jets and gauge bosons).

3.1 Large Hadron Collider

The LHC was constructed between 2000 and 2008 in the same circular tunnel used to house the LEP (Large Electron-Positron Collider), which is 27 km long and 3.8 meter wide, and lies as deep as 175 meters beneath the ground. The overall layout of the LHC is shown in Fig. 3.1.

The nominal collision energy at the LHC is 14 TeV of centre of mass energy ($\sqrt{s}$), which is much higher than any predecessor collider. In 2011, LHC was running with $\sqrt{s} = 7$ TeV, and in 2012, $\sqrt{s} = 8$ TeV. These two-year runs are belonging to the "RUN I" stage of the LHC. In the "RUN II" stage, which is a three-year run and starts from early 2015, it is anticipated to reach a centre of mass energy of first 13 TeV and later 14 TeV.

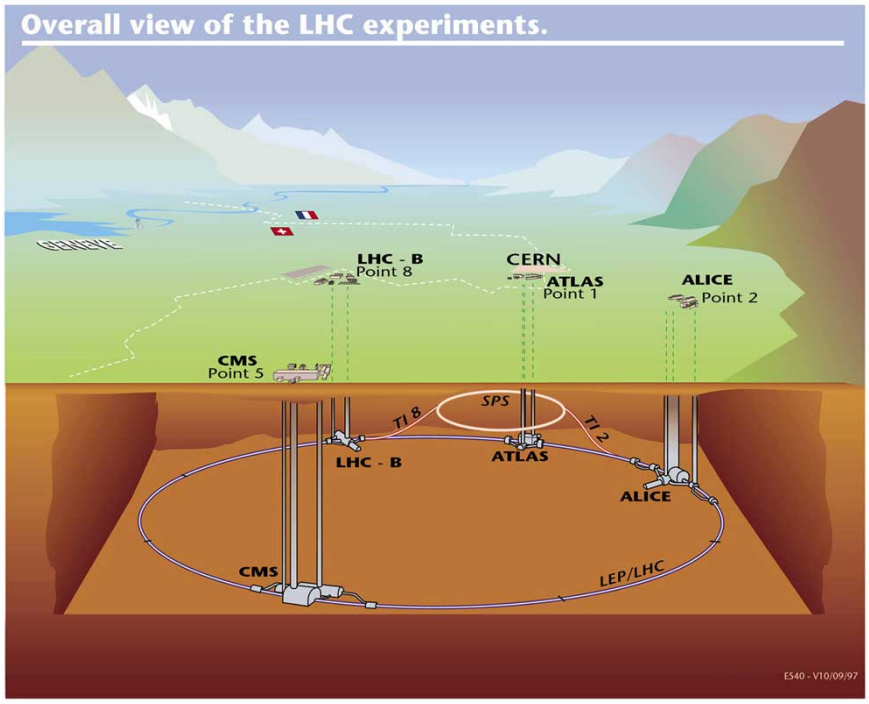


Figure 3.1: The layout of the LHC, which crosses the boarder between France and Swiss. The larger ring is the tunnel for the LHC, which used to be the LEP. The smaller ring aside represents one of the pre-accelerator, the SPS. The locations of the four main detectors are denoted in bold on the map.

3.1.1 Purpose

There are a number of reasons why the LHC was built despite its cost.

First of all, as mentioned in the previous chapter, the Higgs mechanism is very important in the SM for giving particles masses. Whether it is just a mathematic trick or truly the choice of the nature, which results in the Higgs boson, is very important to our understanding of the universe. Before the LHC, a lot of efforts has been made to search for the Higgs boson within the available energy region of the existing accelerators and all attempts failed. Thus we need a more powerful machine to help answering this basic question. Now we know the Higgs is discovered at the LHC, which proved that the investment is worthwhile.

Second, there are alternative theories that are competing with the SM, like the supersymmetry. It is vital to know if they are correct. By looking for particles predicted by those theories, which can only be produced in the energy scale of the LHC, we can validate or rule out them.

Third, since the LHC is a very high energy proton-proton collider, abundant

events from SM processes are produced. With these collision data, precision measurements can be made to further validate the success of the SM or search for deviations from the prediction of the SM.

Finally, the LHC could help to answer some other open questions that are beyond the capability of the SM, like: the existence of extra dimension, the asymmetry between matter and anti-matter, the possibility of the GUT etc.

3.1.2 Design

In its main operation mode and running at full design power of 7 TeV per beam, the LHC collides synchronously two beams of protons (in two different pipes), in the form of tiny bunches, in opposite directions at several interaction points (IPs) of the tunnel where the detectors are installed.

The tunnel is kept to be nearly in vacuum to reduce contamination from the air. And it is in an extremely low temperature (maintained by tonnes of superfluid helium 4) of $1.9K$ (-271.25^oC), which is the operation point of the superconducting magnets that provides the $\sim 8\,\mathrm{T}$ magnetic field to bend or focus the beams and keep them on their circular path. This is not easy since protons are so small that making them collide is akin to firing two needles 10 km apart with such precision that they meet halfway.

At the full design power, protons will have a Lorentz factor of about $7,500$ and move at about $0.999999991c$, or about 3 meters per second slower than the speed of light (c). However, protons are not accelerated to this speed directly. First of all, they pass through the linear particle accelerator LINAC2 to gain a 50-MeV energy. Then they are fed into the Proton Synchrotron Booster (PSB) to reach the energy of 1.4 GeV. And then they enter into the Proton Synchrotron (PS) where they are accelerated to 26 GeV. Finally their energy reaches 450 GeV by the Super Proton Synchrotron (SPS), which is the smaller ring shown in Fig. 3.1. After all these pre-accelerations, the protons are injected into the LHC main ring and accelerated to the final designed energy.

Rather than colliding protons one by one each time (technically extremely hard), they are formed into bunches (115 billion protons per bunch) and it is these bunches (there are in total 2808 of them in the tunnel) that collide each time with a time interval down to 25 nanoseconds. And there will be in average around 20 proton-proton interactions per bunch crossing. This collision rate can be quantified by the so-called "luminosity" (this quantity will be defined in later section) whose value

is expected to be $10^{34} cm^{-2} s^{-1}$ at the full design power.

Besides proton-proton collision, the LHC is also arranged to run in another mode, where heavy-ions are used to replace the protons, which is used to investigate quark-gluon plasma.

3.1.3 Detectors

There are four main detectors at the LHC located underground at different points of the ring. They are the ATLAS, CMS, ALICE and LHCb, whose locations are show in Fig. 3.1.

The ATLAS will be introduced in detail in the following section. The CMS stands for Compact Muon Solenoid: the "Solenoid" comes from the huge solenoid magnet around which the CMS is built. Together with the ATLAS, they are the two general-purpose particle detectors. It locates in the opposite point with respect to the ATLAS, which can be seen from Fig. 3.1. And it uses different technical solutions and design of magnet systems from the ATLAS. However, their sub-detector configurations are the same and they share similar physical goals and are designed to complement each other.

The ALICE means "A Large Ion Collider Experiment". Unlike the ATLAS and CMS, it is dedicated to study a form of matter called quark-gluon plasma by using heavy-ion (Pb-Pb nuclei) collisions. The quark-gluon plasma is a state of matter wherein quarks and gluons are freed (because of the high energy density when the collision happens), which are believed to exist shortly after the Big Bang. This study is important for our understanding, in Quantum ChromoDynamics (QCD), of color confinement and chiral symmetry restoration.

The LHCb is the short for "Large Hadron Collider beauty". As the name suggests, it is a specialised b-physics experiment. By studying the interactions of the b-hadrons, the parameters of CP violation could be measured, which can help to explain the matter-antimatter asymmetry of the universe. Besides, the detector can also be used to measure the production cross sections and electroweak physics.

Except for the four main detectors, there are three more, which are very much smaller and for very specialised research:

- The TOTal Elastic and diffractive cross section Measurement (TOTEM), which is used to measure the total cross section, elastic scattering, and diffractive processes. What to be noticed is that is shares the same interaction point with the CMS,

- The MoEDAL experiment, whose fullname is "Monopole and Exotics Detector At the LHC", is to directly search for the magnetic monopole or dyon and other highly ionising stable massive particles and pseudo-stable massive particles. It shares the same cavern with the LHCb,
- The LHCf, with "f" standing for "forward", uses particles thrown forward by LHC collisions to simulate cosmic rays. Since there are two forward regions, it consists of two detectors, 140 m apart from either side of the interaction point of the ATLAS experiment.

3.2 Proton-proton collision

After the introduction of the LHC machine, the physics processes happened during the proton-proton collision will be described in the following, especially the Higgs production and decay, which is relevant to our study.

3.2.1 Parton distribution functions

Protons are composed of three valence quarks (*up*, *up*, *down*), the gluons, and transitory pairs of sea quarks, all of which are called partons. Thus when two protons collide to each other, it is essentially the partons inside the protons that are taking part in the collision.

Each parton randomly carries a fraction of the proton's momentum and the distribution of that fraction is described by the so-called parton distribution function (PDF) $f(x, Q^2)$, which is defined as the probability density of finding a parton with momentum fraction x at momentum transfer (or energy scale) Q^2.

QCD provides quantitative predictions about how the PDF evolves as the energy scale Q^2 varies by the so-called 'DGLAP' equation [12–15]. However, the x dependence of the PDF at a given Q^2 can not be predicted but has to be measured experimentally.

Thus to determine the PDF from measurements, a parametrization is assumed to be valid at some starting $Q^2 = Q_0^2$. Then the DGLAP equation is used to evolve the PDF to a different Q^2 where predictions of measured quantities are obtained. The predictions are then fitted to the measured datasets, thus constraining the parameters. There are various groups performing fits to the available datasets and providing PDF sets for the proton: for instance the groups ABKM [16], CTEQ [17], HERAPDF [18, 19], MSTW [20], and NNPDF [21].

In Fig. 3.2 [22], two PDFs at two different energy scale Q^2 determined by MSTW are shown. It can be seen that gluons dominant the low x region and the contributions from sea quarks become more important at high Q^2.

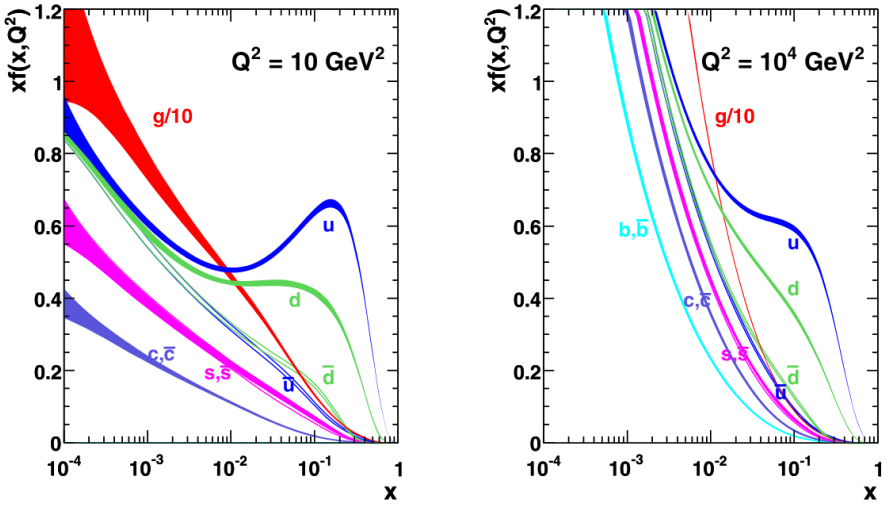


Figure 3.2: PDFs of the proton determined from the MSTW08 PDF set for (left) $Q^2 = 10\,\mathrm{GeV}^2$ and (right) $Q^2 = 10^4\,\mathrm{GeV}^2$. The bands reflect the uncertainties at the 68% confidence level.

3.2.2 Hard and soft processes

Depending on the magnitude of the momentum transfer, the scattering processes between partons can be classified as either hard or soft. For hard processes in which we are usually interested, e.g. high-p_T jet production or W and Z production, the rates and event properties can be predicted with good precision using perturbation theory. For soft processes, e.g. the underlying event, the rates and properties are dominated by non-perturbative QCD effects, which are less well understood.

One typical hard-scattering process between two hadrons A and B can be illustrated in Fig. 3.3 [22]. $f_{a/A}(x_a, \mu_F^2)$ and $f_{b/B}(x_b, \mu_F^2)$ are the PDFs of the two partons a and b which initiate the hard process with the momentum fraction x_a and x_b and the energy scale μ_F^2. $\hat{\sigma}$ is the parton-parton cross section of the hard scattering, which can be calculated with perturbative QCD.

Through a summation over all possible parton-parton pairs and an integration over all possible momentum fractions, we can get the hadron-hadron cross section [22]:

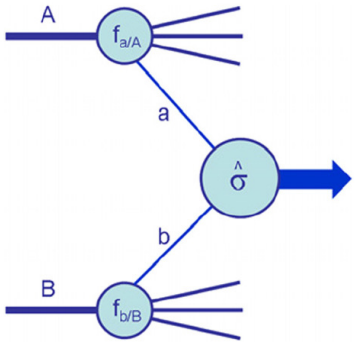


Figure 3.3: A general hard-scattering process.

$$\sigma_{AB} = \sum_{a,b} \int dx_a dx_b f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) \hat{\sigma}_{ab}(x_a, x_b, \alpha_s(\mu_R^2)). \tag{3.1}$$

The partonic level cross section $\hat{\sigma}_{ab}$ depends on the strong coupling constant α_s thus on the renormalisation scale μ_R, which is a point that we choose to define the renormalised quantity when we encounter infinite in practical calculation and want to eliminate this infinite. The physical predictions of the theory, calculated to all orders, should in principle be independent of the choice of renormalisation scale. But in practice, calculations are only done to finite orders, in which case, the uncertainties will be affected by the chosen renormalisation point. So usually μ_R is set to be the energy scale characteristic of the experiment. The scale μ_F in the formula is the factorisation scale, which can be thought of as the scale that separates long- and short-distance physics. Below this scale, perturbation theory is no longer trustable, and the soft and collinear divergences can be absorbed into the PDF. In summary, the renormalisation scale is arbitrary, but it should be chosen wisely so that large logarithms vanish. But the factorisation scale is not arbitrary. It is set by the kinematics of the problem and denotes the point below which perturbative QCD is not accurate any more.

As mentioned, the partonic level cross section $\hat{\sigma}_{ab}$ can be calculated in perturbative QCD to any orders. The leading order (LO) cross section is denoted as $\hat{\sigma}^0_{ab}$, the next-to-leading order (NLO) as $\hat{\sigma}^1_{ab}$, etc. Usually, the higher order cross sections are parametrised in terms of the total K factors, which are defined as the ratio of the cross section computed to the that order to the Born level cross section:

$$\sigma^{[n]} = \sigma^{[0]} \cdot K^{[n]}. \tag{3.2}$$

Those partons in a pp collision which do not take part in the hard scattering process will produce what are generally called "underlying event" (UE), which are soft due to their small momentum transfer. In practical calculations, these events are modelled phenomenologically by the so-called Parton Shower (PS) programs with careful tuning.

What we are interested in is the rare hard process so that bunches containing $\sim 10^{11}$ protons collide in a short time gap of several tens of ns. Usually there will be ~ 20 (or more in Run II) proton-proton collisions in a bunch crossing and there may be one collision where a hard process happens. The rest collisions without hard process are called pile-up, or more accurately in-time pile-up because they are in the same time window as the hard process. There is also out-of-time pile-up, which is a result of read-out time of the calorimeter being larger than the bunch crossing time resulting in the collisions of neighbour bunch crossings being mixed with the hard process.

3.2.3 Example hard processes

At the LHC, the total inelastic proton-proton cross section could reach $\sim 70\,\text{mb}$ (we are not interested here in elastic collisions), of which many important hard processes could be studied either to test the QCD or to search for new particles.

The dominant processes are those mediated by strong interaction, like the dijet events, There will also be a small fraction of electroweak processes, like the W and Z production, whose cross section is about six or seven orders of magnitude smaller than the total inelastic cross section. The Higgs production cross section is expected with to be ten or eleven orders of magnitude smaller than the total inelastic pp cross section [22].

When an incoming parton from one proton scatters off that from the other proton to produce two high-p_T partons, these two partons will form two jets through a mechanism similar to the electromagnetic cascade. This type of process is called dijet production, whose leading order Feynman diagrams are shown in Fig. 3.4 [22].

When a quark and an antiquark from the incoming protons annihilate, the vector bosons W and Z can be produced. The Feynman diagrams in LO and NLO for these processes, are shown in Fig. 3.5 [22].

The Feynman diagrams for the production of pairs of W bosons are shown in Fig. 3.6 [22]. The dominant W^+W^- production mechanisms are s-channel and t-

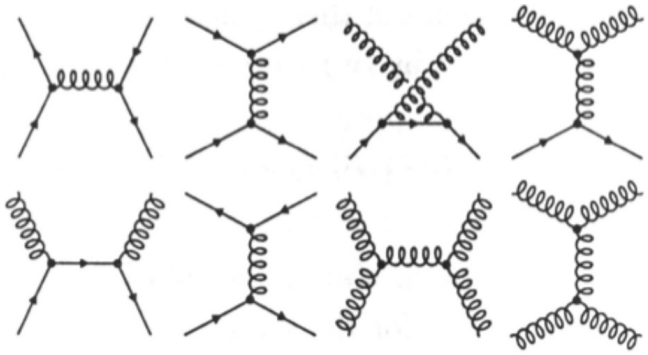


Figure 3.4: Leading order diagrams for the production of dijet events

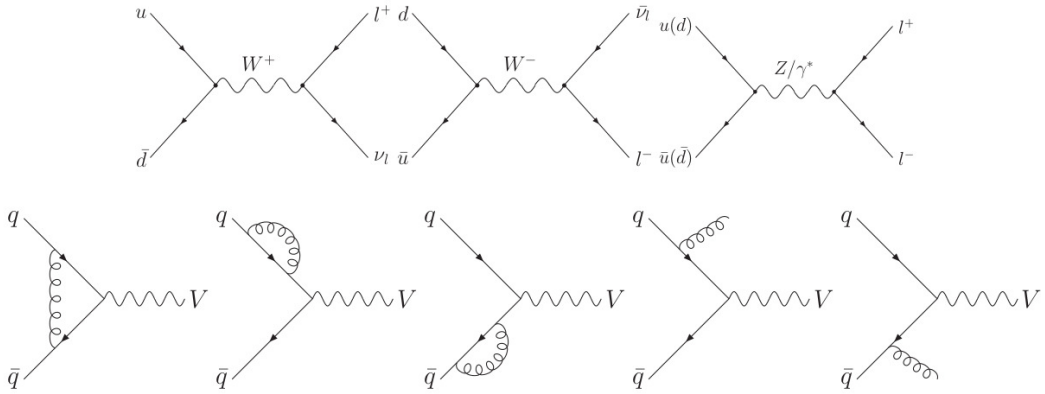


Figure 3.5: Leading order (top) and some next-to-leading order diagrams (bottom) for the production of W and Z bosons

channel quark-antiquark annihilation. This process at the LHC is of great interest since it provides an excellent opportunity to test the predictions on the structure of the gauge couplings of the electroweak sector at the TeV energy scale.

3.2.4 Higgs production

Through proton-proton collisions at the LHC, Higgs bosons can be produced via four different production mechanisms: gluon fusion, vector boson fusion, associated production with weak gauge bosons, and associated production with heavy quarks, whose Feynman diagrams are all shown in Fig. 3.7.

The gluon fusion process is the dominant Higgs production mechanism over a large range of possible masses. Other processes are also of interest due to their special signatures. In Fig. 3.8, the production cross sections of each mechanisms, together with their order of perturbative calculations and uncertainties, are shown over the Higgs mass for a pp collision energy of 8 TeV.

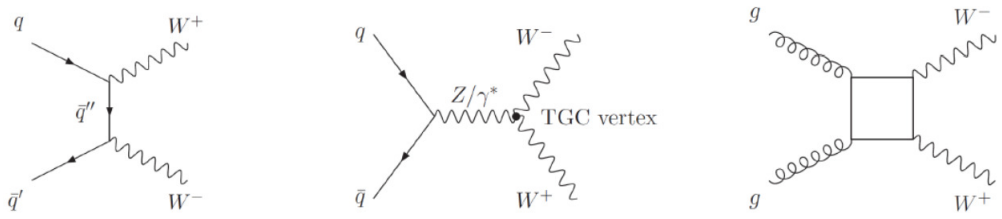


Figure 3.6: (left): The SM tree-level Feynman diagram for W^+W^- production through the $q\bar{q}$ initial state in the t-channel. (Middle): The corresponding SM tree-level diagram in the s-channel, which contains the WWZ and $WW\gamma$ triple gauge boson coupling (TGC) vertices. (Right): The gluon fusion process, mediated by quark loops.

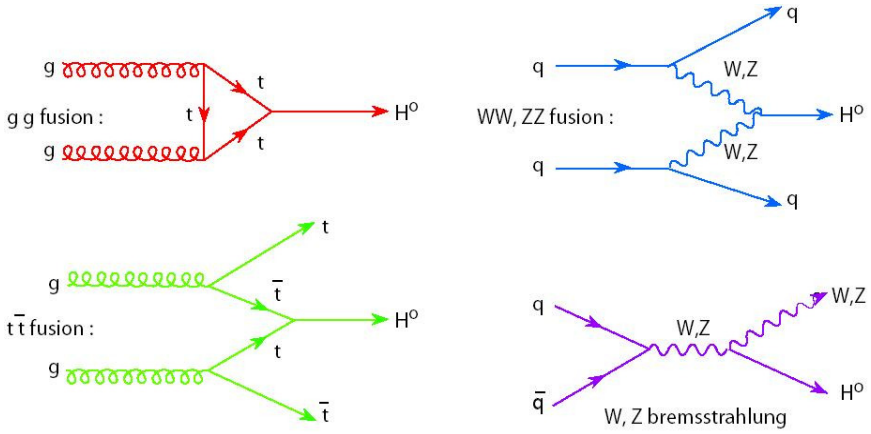


Figure 3.7: Feynman diagrams for the four main types of Higgs boson production.

Gluon gluon fusion

The gluon gluon fusion (ggF) process for the Higgs production $gg \to H$ is shown in red in Fig. 3.7, where the Higgs boson is produced from two initial state gluons mediated by virtual heavy fermions that couple to the Higgs boson. The loop is totally dominated by the top quark (7% by the bottom quark) because of the strong Higgs coupling to the heavy top quark. The partonic level cross section is:

$$\hat{\sigma}(gg \to H) = \frac{8\pi^2 \Gamma_{H\to gg}}{N_g^2 m_H} \delta(\hat{s} - m_H^2), \tag{3.3}$$

where $\Gamma_{H\to gg}$ is the width of the $H \to gg$ decay, $N_g = 8$ is the number of different gluons, m_H the Higgs mass, and $\hat{s} = x_1 x_2 s$ the squared centre of mass energy of the gluon pair.

To get the full cross section, Formula 3.1 is used:

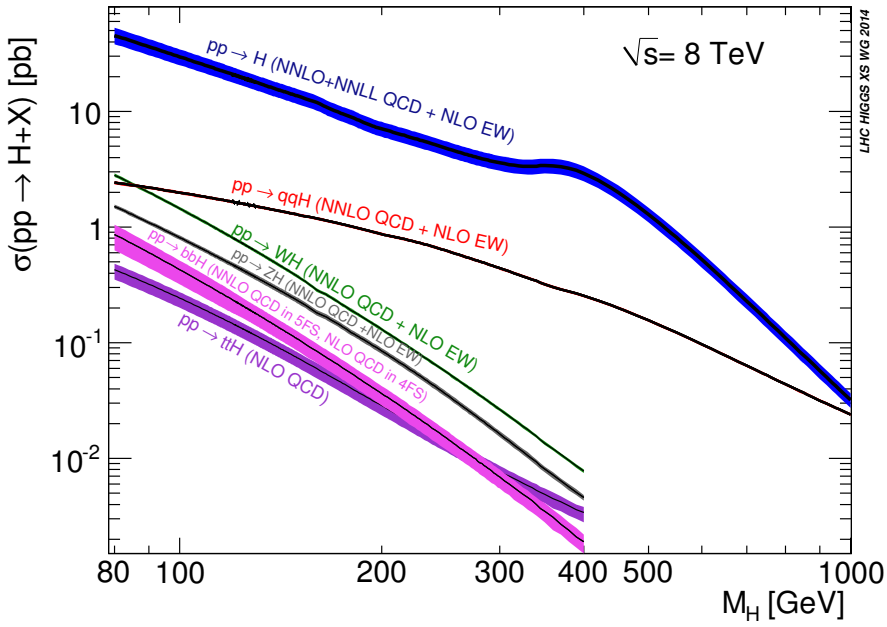


Figure 3.8: Higgs boson production cross section of each mechanism as a function of the Higgs mass. Error bands are shown in color.

$$\sigma_0(pp \to H) = \iint dx_1 dx_2 f(x_1, m_H^2) f(x_2, m_H^2) \hat{\sigma}(gg \to H). \tag{3.4}$$

From the formula above, it is easy to notice that the value of the cross section has two main uncertainties. The first is from the gluon PDF which still has a large uncertainty in the low and high x region. The second is from the partonic level cross section $\hat{\sigma}(gg \to H)$, whose precision is limited by the available perturbative calculation.

Vector boson fusion

The vector boson fusion (VBF) process for the Higgs production $qq \to qqH$ is shown in blue in Fig. 3.7, where the Higgs boson is produced at tree level by two vector bosons, leaving a distinct signature of two high energy jets in the forward and backward direction, with a large gap in pseudorapidity between them.

In the low mass region where the Higgs is found, the VBF process contributes roughly one order of magnitude smaller than the ggF process. But in the high mass region, the VBF process becomes more important since the coupling to longitudinal polarized vector boson is getting stronger. At around 1 TeV, the VBF production cross section reaches the same level as that of the ggF.

Associated production

The Higgs boson can be produced associate with two kinds of particles, the vector bosons and the heavy quarks. The former process is usually called the Higgs Strahlung $qq \to W/Z\,H$ which is shown in purple in Fig. 3.7. In this case, a quark and anti-quark merge to form a virtual W or Z boson which, if it carries sufficient energy, can then emit a Higgs boson. For a Higgs mass of 125 GeV, the production cross section is around 6% of that of the ggF.

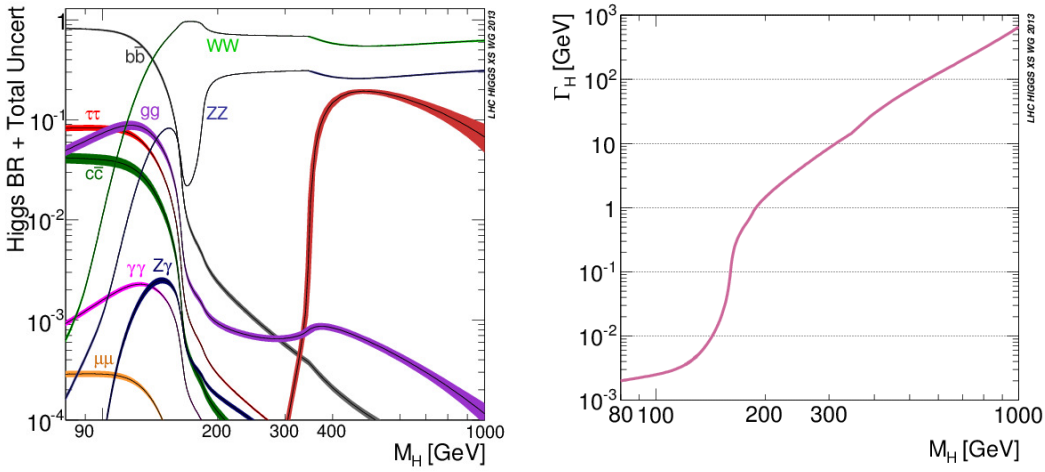
The latter process is mainly dominated by associate production with top quark $gg \to t\bar{t}H$, which is shown in green in Fig. 3.7. In this case, two colliding gluons each decay into a heavy quark-antiquark pair, where a quark and an antiquark from each pair fuse into a Higgs boson. This process has the least cross section among all production mode, e.g. for a Higgs mass of 200 GeV, the cross section is around a factor of five below that of the Higgs Strahlung.

3.2.5 Higgs decay

Since a heavy Higgs boson is expected to interact with all the massive elementary particles, it decays very soon with an average life time in the order of 10^{-22} s. The Higgs boson has many different possibilities to decay, each having its own probability, which is quantified by the so-called branching ratio, defined as the fraction of the total number decays that follows that process.

The branching ratios of the important decay channels together with their total uncertainties are shown in Fig. 3.9(a). The $b\bar{b}$ channel dominates in the low mass region, while the W^+W^- channel is predominant after 150 GeV mass region. And the total decay width of the Higgs boson is shown in Fig. 3.9(b). From the plot, when the Higgs mass is below 170 GeV, there is a narrow resonance peak at each mass point, while above 170 GeV, no clear resonance is expected. It should be noticed is that new physics could change significantly the Higgs width.

The Higgs boson could decay into a fermion-antifermion pair. Since the interaction strength between the fermions and Higgs is proportional to the fermion mass, the heavier the fermions are, the more likely the decay happens. Thus if the Higgs mass would be above twice of the top mass, which is ~ 346 GeV, the top-antitop decay dominants the fermion-antifermion channels. However, if the Higgs mass is below that threshold, the top-antitop mode will be heavily suppressed, and instead the bottom-antibottom pair decay dominants. At the Higgs mass point ~ 125 GeV, the Higgs decays into a $b\bar{b}$ pair with probability of 57.7% and the second dominant



(a) (b)

Figure 3.9: Higgs decay branching ratio (a) and Higgs decay width (b) as a function of Higgs mass.

decay channel at that mass point is the Higgs to tau-antitau pair, of which the probability is 6.3%.

The Higgs can also split into a pair of massive gauge bosons with one of them being off-shell. The most probable one is the Higgs decaying into W^+W^-, which is the leading decay channel over a large mass range and especially has a probability of 21.5% at Higgs mass point of 125 GeV. The outgoing W bosons can decay subsequently into a quark and an antiquark or into a charged lepton and a neutrino. In the former case, the decay is hard to distinguish from the large QCD backgrounds. In the latter case, although there is clear signature of high-p_T leptons in the final state, the neutrinos are impossible to detect, resulting in the insensitivity of the $H \to WW^* \to \ell\nu\ell\nu$ channel to the Higgs mass. It is this channel that will be used in the following study of the Higgs observation and its property measurements. The other important decay channel is the Higgs to a pair of Z bosons if each of the bosons subsequently decays into a pair of electrons or muons, which is easy to detect. This channel is cleaner than the W^+W^- channel, However, its branching ratio is a much smaller than that of the W^+W^- channel, e.g. at Higgs mass point of 125 GeV, the Higgs decays into Z-bosons with a probability of only 2.6%.

All the above decays happen at tree level, however, the Higgs boson could decay into a pair of massless gauge bosons through virtual loop of heavy quarks or massive gauge bosons. The most common such process is the Higgs decaying into a pair of gluons through a loop of virtual heavy quarks, which is the reverse of the ggF

production of the Higgs boson. At mass point of 125 GeV, the probability of this decay mode is 8.6%. Another possibility is the decay into a pair of photons by a loop of W bosons or heavy quarks, which has a probability below 1%. However, this process is very important for the Higgs boson search since the energy and momentum of the photons could be measured very precisely, thus reconstruction of the decaying particle could be done accurately making it a perfect channel for the Higgs mass measurement.

As mentioned before, among all the decay modes introduced before or presented in the plots, the W^+W^-, ZZ, and $\gamma\gamma$ are the most important ones since they three constitute the discovery of the Higgs boson in 2012.

3.3 ATLAS detector

We are going to introduce the ATLAS detector in this section, the contents of which are from the ATLAS Technical Design Report (TDR) [23]. The overall layout of the detector is shown in Fig. 3.10. It is 46 metres long, 25 metres in diameter and 7000 tonnes in weight. From inside to outside, it is the Inner Detector (ID), the calorimeter and the Muon Spectrometer (MS).

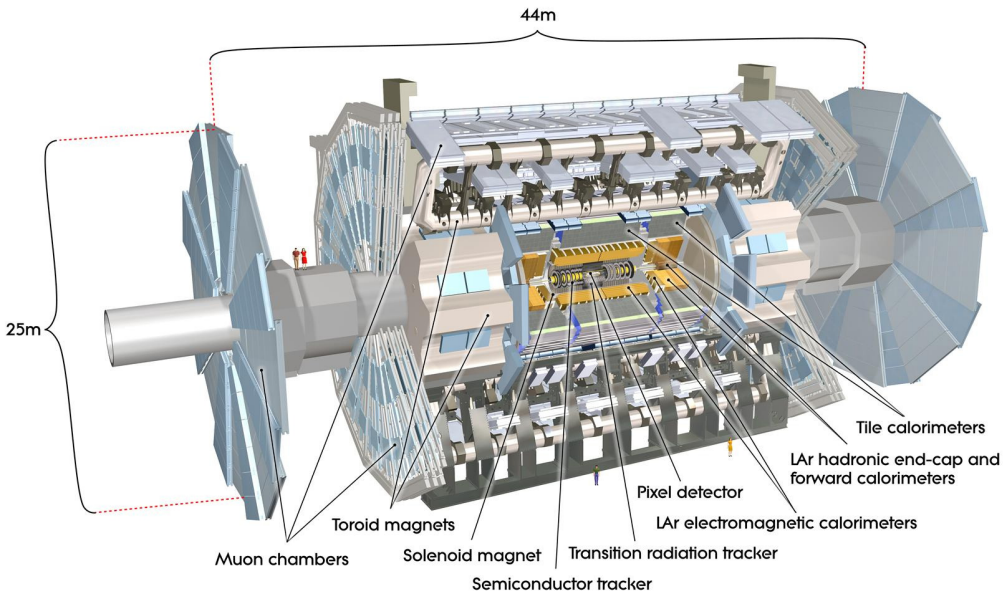


Figure 3.10: The layout of the ATLAS detector.

The ID, which is the tracking volume, consists of three sub-detectors: the semi-

conductor pixel and strip detectors in the inner part and the continuous straw-tube tracking detectors with transition radiation capability in its outer part. The calorimeters are divided into two types based on its purpose: the electromagnetic (EM) calorimeter in front and the hadronic calorimeter behind, which have different space coverages and apply different technologies. The MS, which defines the overall dimensions of the ATLAS detector, is composed of three stations of high-precision tracking chambers, where four different chamber technologies are implemented.

Between the ID and the calorimeter, it is the superconducting solenoid magnet surrounding the ID cavity, which provides bending magnetic fields to the tracking volume. Outside the calorimeter, the air-core toroid superconducting magnet arranged with an eight-fold symmetry provides strong magnetic fields for the measurements in the MS.

ATLAS is a general purpose detector which is used to investigate many different types of physics: Some of them are confirmations or improved measurements of the SM, while many others are possible clues for new physical theories.

In the following sections, each component of the ATLAS detector will be described specifically.

3.3.1 Nomenclature

Before introducing the ATLAS detector component by component, the following nomenclature will be defined since they will be used in following sections to better describe the detector.

The coordinate system is defined with the z-axis being the beam direction and the $x-y$ plane being the transverse plane with respect to the z-axis. The positive x-axis is pointing from the interaction point to the centre of the LHC ring and the positive y-axis is pointing upwards. The azimuthal angle ϕ is measured around the beam axis and the polar angle θ is the angle from the beam axis. The pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$. The transverse momentum p_T and the transverse energy E_T, as well as the missing transverse energy E_T^{miss} and other transverse quantities, are defined in the $x-y$ plane. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Besides, the $R\phi$ direction will be defined as the tangent of the circle of radius R in the transverse plane (the R here is to be distinguished from the one in ΔR).

3.3.2 Magnet system

The ATLAS superconducting magnet system can be seen in Fig. 3.11, which determines the size of the ATLAS to be 46 m in length and 25 m in diameter. It is composed of one solenoid and two toroids providing strong bending power for the measurement of the track of the charged particle. These large superconductor magnets are indirectly cooled by forced flow of helium at 4.5 K and powered by several or several tens of kA power supply.

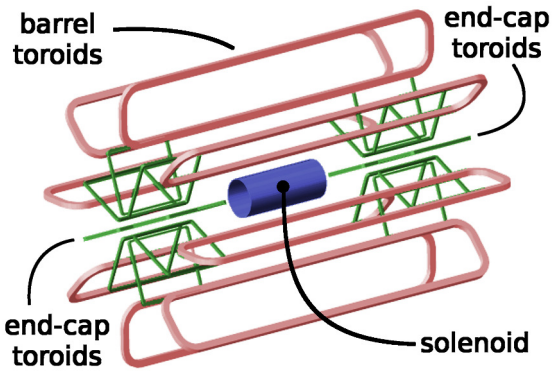


Figure 3.11: The magnetic systems of the ATLAS detector.

The solenoid in the centre provides a magnetic field of 2 T for the ID in the beam direction. It is in cylinder shape and hollow, with dimensions of 5.3 m in length and 2.4 m in diameter, and located in front of the electromagnetic calorimeter.

The three large air-core toroids, one in the barrel and two in the end-cap, generate magnetic field for the muon spectrometer. Each of them consists of eight coils assembled radially and symmetrically around the beam axis. The two end-cap toroids are inserted with a rotation of 22.5° in the barrel toroids at each end and line up with the solenoid. The performance is characterised by the field integral $\int Bdl$, where B is the azimuthal field component and the integral is taken on a straight line between inner and outer radius of the toroids. The barrel toroids provides 2 to 6 Tm and the end-cap toroids provides 4 to 8 Tm. The bending power is lower in the transition region where the two magnets overlap.

3.3.3 Inner detector

The layout of the ID is shown in Fig. 3.12. The outer radius of the ID is 115 cm and the total length is 7 m. It consists of high-resolution pixel detector in the inner radii, silicon microstrip (SCT) detector in the intermediate radial range and

continuous straw tube tracker (TRT) at the outer radii, all contained in the central solenoid introduced before and all composed of one barrel unit and two end-cap units, which provide full tracking coverage over $|\eta| \leq 2.5$.

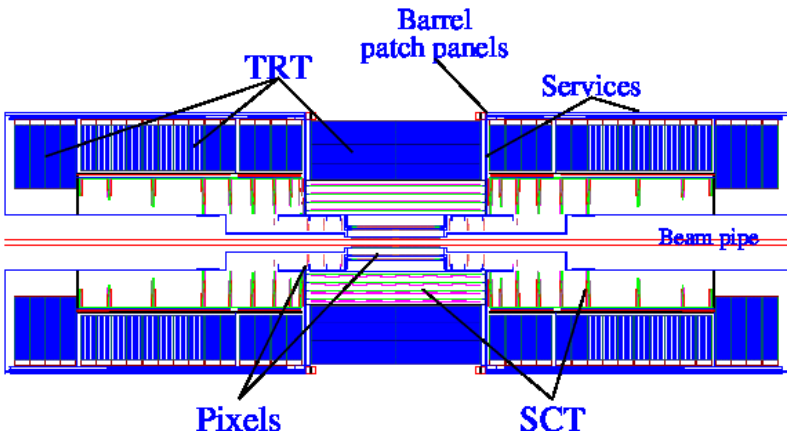


Figure 3.12: The Inner Detector of the ATLAS detector.

The pixel system consists of three barrels at average radii of 4 cm, 10 cm, and 13 cm, and five disks on each side, between radii of 11 and 20 cm, which complete the η coverage. The 140 million detector elements, each 50 μm in $R\phi$ direction and 300 μm in z, are organised into modules (each module contains 61 440 elements) and distributed over these layers and disks. Such a design makes the pixel system be able to provide a very high-granularity, high-precision set of measurements as close to the interaction point as possible. In average, three precision measurements will be provided by the system over the full acceptance, which mostly determines the impact parameter resolution and the ability to find short-lived particles such as B hadrons and τ leptons. The expected resolution will be 12 μm in $R\phi$ and 66/77 μm in z for barrel/end-cap.

The SCT, which is the abbreviation for SemiConductor Tracker) consists of four barrels at radii of 30, 37, 45 and 52 cm, and nine end-cap wheels on each side, between radii of 27 and 56 cm, which are also modularised. A single SCT unit is $6.36 \times 6.40\,\text{cm}^2$, which is almost a square, with 768 strips of 80 μm pitch. Two of such units are bonded together to form a 12.8 cm (sometimes 6–7 cm at the innermost end-cap wheel) long strip and two such strips are glued together back-to-back at a 40 mrad angle to form a SCT module. Eight precision measurements will be provided by the SCT when it is crossed by a charged particle, giving four space points with resolution of 16 μm in $R\phi$ and 580 μm in z/R for barrel/end-cap layers.

The TRT is made of bunches of straw detectors, which is a 4 mm diameter tube with a 30 μm diameter gold-plated W-Re wire in the centre and fulfilled with a non-flammable gas mixture of Xe, CO_2 and CF_4. In the barrel section, straws are in parallel with beam axis and there are 73 layers of them covering the radial range from 56 to 107 cm. While in each end-caps, straws are perpendicular to the beam axis and there are 160 layers of them installed on 18 wheels which cover the radial range from 48 to 103 cm. The gaps between the straws are filled with a radiator that will causes transition radiation if an electron passes through, which can be used to do electron identification. The design of the TRT gives in average 36 measurements per track and 170 μm resolution per straw. The low resolution is compensated by the large number of measurements so that the final resolution of the TRT could compete with that of silicon detector.

3.3.4 Calorimeters

A view of the ATLAS calorimeters is presented in Fig. 3.13, which consists of an EM calorimeter covering the pseudorapidity region $|\eta| < 3.2$, a hadronic barrel calorimeter covering $|\eta| < 1.7$, two hadronic end-cap calorimeters covering $1.5 < |\eta| < 3.2$, and two forward colorimeters covering $3.1 < |\eta| < 4.9$.

These different sub-calorimeters cover different η range, use different techniques, and are constructed with different granularity. The granularity is quantified by the size of the calorimeter cell, which is the basic unit of the calorimeters. The calorimeters are usually segmented into several layers and the cell size, measured in $\Delta\eta \times \Delta\phi$, differs in different layers.

The EM calorimeter, which is mainly used to measure the energy of electrons and photons, is divided into a barrel part ($|\eta| < 1.475$) and two end-cap parts ($1.375 < |\eta| < 3.2$). The barrel calorimeter consists of two identical half-barrels, separated by a small gap (6 mm) at $z = 0$. Each end-cap calorimeters is mechanically divided into two coaxial wheels: an outer wheel covering the region $1.375 < |\eta| < 2.5$, and an inner wheel covering the region $2.5 < |\eta| < 3.2$. It is a lead LAr detector with accordian-shaped Kapton electrodes and lead absorber plates, which in total corresponds to > 24 radiation lengths (X_0) in the barrel and $> 26X_0$ in the end-caps. In Fig. 3.14, the typical energy resolutions of the electrons and photons are shown as a function of the incident energy, which are of order $10\%/\sqrt{E(\mathrm{GeV})}$.

The ATLAS hadronic calorimeters use different techniques in different pseudo-

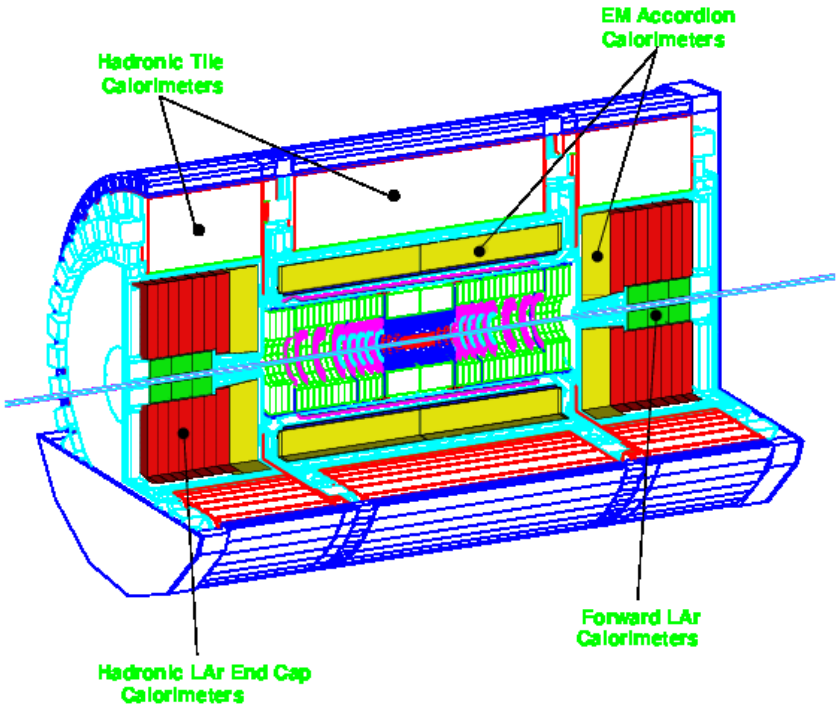
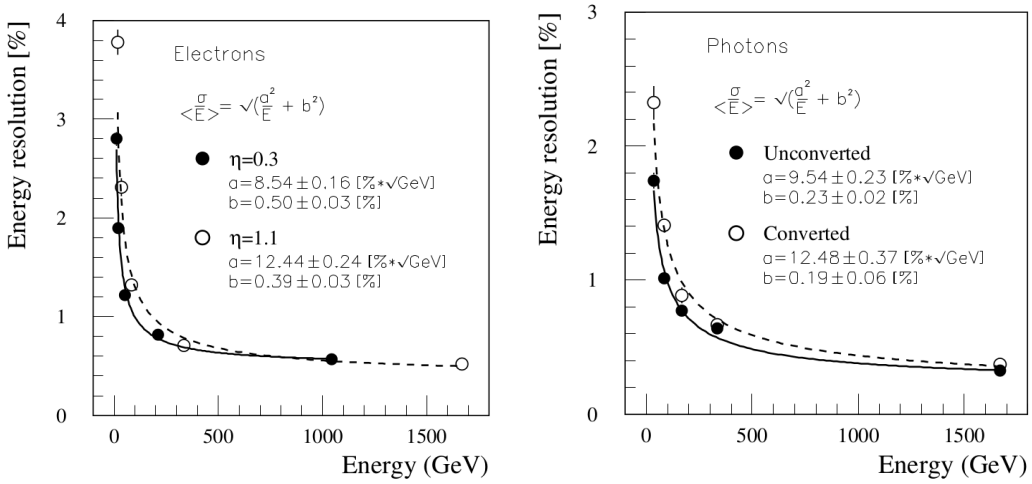


Figure 3.13: The calorimeters of the ATLAS detector.



(a) (b)

Figure 3.14: The energy resolution (a) for electrons at $|\eta| = 0.3$ and $|\eta| = 1.1$, and (b) for converted or unconverted photons at $|\eta| = 1.1$, as a function of the incident energy.

rapidity ranges. Over the range $|\eta| < 1.7$, the iron scintillating-tile technique is used for the barrel and extended barrel tile calorimeters. Over the range $1.5 < |\eta| < 4.9$, the LAr calorimeters were chosen: the hadronic end-cap calorimeter (HEC) extends to $|\eta| < 3.2$, while the range $3.1 < |\eta| < 4.9$ is covered by the high-density forward calorimeter (FCAL). The tile calorimeter, which is composed of one barrel and two extended barrels, is a sampling calorimeter using iron as absorber and scintillating tiles as the active material, and extends from an inner radius of 2.28 m to an outer radius of 4.25 m. The HEC and FCAL use the same sensitive material as the EM calorimeter, which is the LAr. The HEC consists of two wheels of outer radius 2.03 m, which are built out of 25/50 mm copper plates. While the FCAL is placed 4.7 m from the interaction point, and it is a particularly challenging detector owing to the high level radiation it has to cope with (because it is close to the beam direction).

3.3.5 Muon spectrometer

The layout of the muon spectrometer (MS) is visible in Fig. 3.15. It is based on the magnetic deflection of muon tracks in the large superconducting air-core toroid magnets, thus it has a similar size as the toroids.

Muon tracks are measured in chambers which are arranged in three stations both in the barrel covering $|\eta| < 1$ and end-caps covering $1 < |\eta| < 2.7$. Not all chambers are the same. Four different techniques are used, which are the Monitored Drift Tubes (MDTs), the Cathode Strip Chambers (CSCs), the Resistive Plate Chambers (RPCs), and the Thin Gap Chambers (TGCs). The former two are precision chambers while the latter two are fast response chambers.

The MDTs are built of aluminium tubes of 30 mm diameter and 400 μm wall thickness, with a 50 μm diameter central W-Re wire. The single-wire resolution is $\sim 80\,\mu$m. But the MDTs contain two multi-layers (three or four layers) of such tubes, thus improving the resolution to $\sim 30\,\mu$m. The MDTs are used in most pseudorapidity ranges (both in barrel and end-caps).

The CSCs are multiwire proportional chambers with cathode strip readout and with a symmetric cell in which the anode-cathode spacing is equal to the anode wire pitch. It has a position resolution of better than 60 μm. Besides, there are some other important characteristics like small electron drift time (30 ns), good time resolution (7 ns), good two-track resolution, and low neutron sensitivity. The CSCs are used only in the end-cap station closest to the interaction point, covering

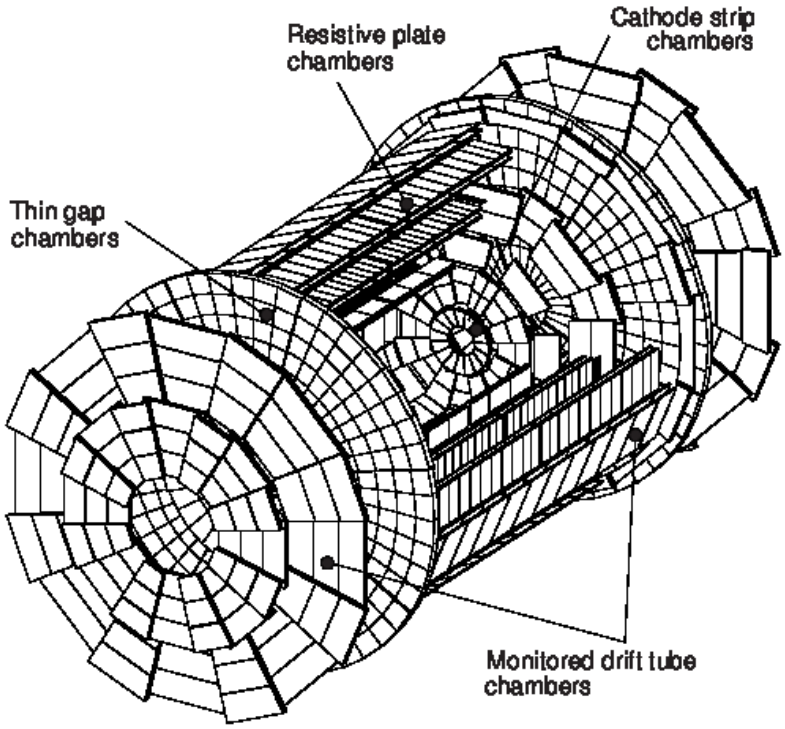


Figure 3.15: The muon spectrometer of the ATLAS detector.

the range $2.0 < |\eta| < 2.7$.

The RPC is a gaseous detector providing a typical space-time resolution of $1\,\text{cm} \times 1\,\text{ns}$. Its basic unit is a narrow gas gap formed by two parallel resistive plates. A RPC chamber is made from two rectangular detector layers, each one read out by two orthogonal strips: the 'η strips' are parallel to the MDT wires and provide the bending view of the trigger detector; the 'ϕ strips', orthogonal to the MDT wires, provide the second-coordinate measurement.

The TGCs are similar in design to multiwire proportional chambers, with the difference that the anode wire pitch is larger than the cathode-anode distance. Signals from the anode wires, arranged parallel to the MDT wires, provide the trigger information together with read-out strips arranged orthogonal to the wires. These wires are also used to measure the second coordinate.

The RPCs and TGCs are fast response chambers: the former are used in the barrel and latter the end-caps. They play a threefold purpose: bunch crossing identification (they have a several nanoseconds time resolution, much smaller than the LHC bunch spacing of 25 ns), trigger (with well-defined p_T cut-offs), and mea-

surement of the second coordinate in a direction orthogonal to that measured by the precision chambers.

3.3.6 Trigger and data-acquisition system

The ATLAS trigger and data-acquisition (DAQ) system is used to select the very rare processes, of main interest to us, out of the very large amount of collisions ($\sim 10^9$ Hz at designed luminosity). Trigger, as the name suggests, is used to trigger the storage of the event if the designed criteria are satisfied, e.g. the appearance of a high-p_T electron/muon/jet. It is based on three levels of online event selection, which is shown in Fig. 3.16. Each trigger level refines the decisions at the previous level and, where necessary, applies additional selection criteria (thus has an increasing latency and more time to make a decision).

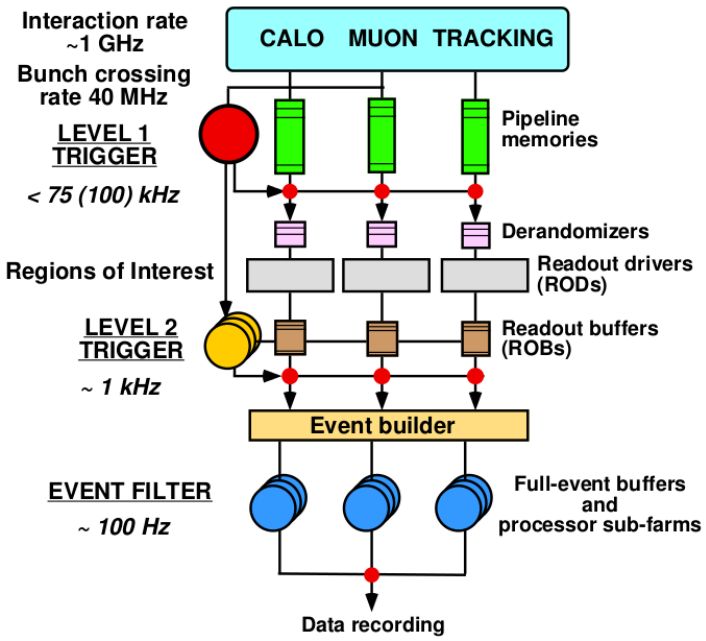


Figure 3.16: The block diagram of the Trigger/DAQ system.

The level-1 trigger (LVL1) is a hardware based system that reduces the event rate from 40 MHz to 75 kHz within a latency of less than 2.5 μs. It is composed of a Calorimeter Trigger (L1Calo), a Muon Trigger (L1Muon) and the event-decision part called Central Trigger Processor (CTP). The L1Calo is based on reduced-granularity information from all the calorimeters and searches for high-p_T electrons and photons, jets, and hadronic decaying τ-leptons, as well as large missing and total transverse energies. The L1Muon uses trigger chambers (RPCs and TGCs)

to look for high-p_T muons. The CTP combines information from the L1Calo and L1Muon to make the trigger decision. Besides, the LVL1 performs the bunch-crossing identification.

Events accepted by LVL1 are passed to the level-2 trigger (LVL2), the 'region-of-interest' (RoI) information provided by the LVL1 is used, including information on the position (η and ϕ) and p_T of the candidate objects, and energy sums. The LVL2 usually only needs to use a few percent of the full event data, corresponding to limited regions centred on the objects indicated by the LV1. The LVL2 will reduce the event rate to $\sim 1\,\text{kHz}$ with a latency to in the range of 1–10 ms. After event passing LVL2, data are transferred by the DAQ system from the Read Out Buffers (ROBs) to the Event Filter (EF) which will be used for the third level event selection.

The EF, which is the last stage of the online selection, together with the LVL2 form the High Level Trigger (HLT) system. It will employ offline algorithms and methods, adapted to the online environment, and use the most up to date calibration and alignment information and the magnetic field map. The output rate from EF should then be reduced to $\sim 100\,\text{Hz}$, corresponding to an output data rate of $\sim 100\,\text{MB/s}$.

3.4 Event reconstruction

The output of the ATLAS trigger and DAQ system is a set of *ByteStream* or *RAW* files, in which the information of the detector response, such as times and voltages, to the particles pass through them are stored, Before being analysed, these raw informations are processed to reconstruct basic quantities like vertices, tracks, and clusters. These quantities are combined to obtain the final physical objects to be used in the analysis, such as photons, electrons, muons, jets, b-jets and E_T^{miss}.

3.4.1 Track

When a charged particle flies through the ID or the MS, its interaction with the detector material is measured as hit information which is then used to reconstruct the particle's trajectory.

The tracks in the ID are reconstructed using position measurements from the ID through several steps:

- First of all, pre-processing algorithms generate silicon clusters, drift circles and space points as input to the pattern recognition,
- Then the track finder starts in the Pixel+SCT detectors and builds the silicon track candidate using pattern recognition,
- Finally, the silicon track is extended to the TRT and a full track fit is performed to determine the track parameters.

Besides, there is another complementary strategy which starts from the TRT segments and reconstructs the track in an outside-in style. It is mainly aimed at reconstruction of photon conversions in the detector and decay vertices of neutral particles.

The reconstructed tracks are described by five helix parameters in an ideal uniform magnetic field with all quantities measured at the point of closest approach to the nominal beam axis $x = 0$, $y = 0$. Parameters in $x - y$ plane are:

- $1/p_T$, Reciprocal of the transverse momentum with respect to the beam-axis,
- ϕ, Azimuthal angle, where $\tan\phi = p_y/p_x$,
- d_0, Transverse impact parameter, defined as the transverse distance to the beam axis at the point of closest approach.

Parameters in the $R - z$ plane are:

- $\cot\theta$, Cotangent of the polar angle, where $\cot\theta = p_z/p_T$,
- z_0, Longitudinal impact parameter, defined as the z position of the track at the point of closest approach.

As for the tracks in the muon spectrometer, they are reconstructed with a similar strategy as in the ID.

3.4.2 Vertex

Although the LHC beam spot is already very small with $\sigma_x = \sigma_y = 15\,\mu$m and $\sigma_z = 56\,$mm, this precision is still not enough for the identification of b- and τ-jets which is important for some physics analysis. However, we can use the precise track parameters determined from silicon trackers to extrapolate the tracks back to the interaction region and look for their probable points of origin with high precision.

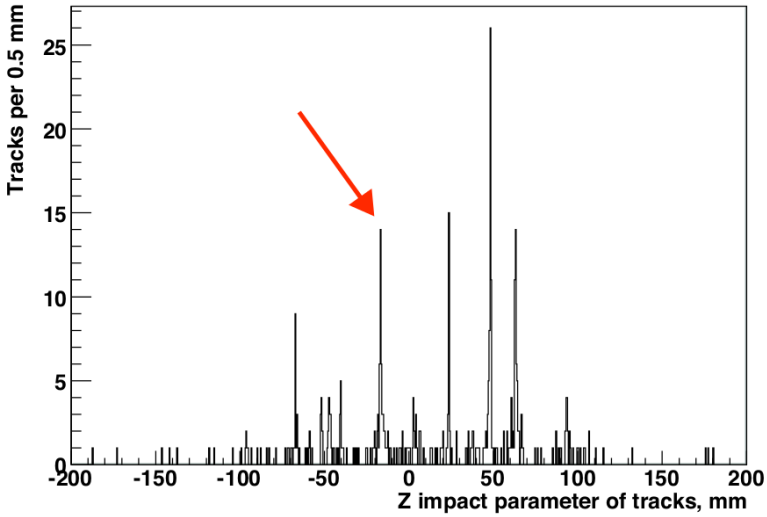


Figure 3.17: Distribution of the z impact parameter of reconstructed tracks in a simulated $H(130) \to \gamma\gamma$ event at $L = 2 \cdot 10^{33} cm^{-2} s^{-1}$. The indicated cluster of tracks corresponds to the signal vertex.

For example, in Fig. 3.17, a simple clustering of the z impact parameters of the reconstructed tracks could help to separate the different vertices in one event.

The primary vertex reconstruction is very important since it is usually the point where the hard process happens. There are two different algorithms for primary vertex finding in ATLAS. One is a simple "fitting after finding" method, which works by clustering pre-selected tracks in the z-projection to determine the number of primary vertices. Then it reconstructs them using vertex fitter. The other one is a "finding through fitting" method and is called Adaptive Multi-Vertex Fitter. It starts with a single seed and increases the number of seeds by forming new ones out of outliers from the fit to the existing vertices. Then an iterative annealing procedure is used during the simultaneous fit of several vertices, such that a hard track-to-vertex assignment is approached. This method gives best performance in terms of both efficiency and precision.

For the secondary vertex reconstruction, ATLAS uses a combination of two algorithms: one forms a discriminator based on the impact parameters of displaced tracks, the other exploits the properties of explicitly reconstructed b-decay vertices. Besides, there is a relatively new algorithm called JetFitter, which creates a tagging likelihood for the reconstructed vertices based on the invariant mass, energy fraction and flight length significance.

3.4.3 Calorimeter cluster

When an incoming particles pass through the calorimeters, their energies are deposited in many calorimeter cells, both in the lateral and longitudinal directions. The energies in these cells are reconstructed on a basic EM energy scale, which has been determined for each module in the system using electron test beam and simulations [24]. These cells are then grouped together into clusters by dedicated clustering algorithms and a total deposited energy of all the cells inside a cluster is assigned to that cluster. These energies are further calibrated, which are different for electrons, photons, and jets, to account for the energy deposited outside the cluster and in dead material.

There are two types of clustering algorithms. The first is called "sliding-window" algorithm, which is based on summing cells within a fixed-size rectangular window. The algorithm keeps adjusting the position of the window so that the total energy deposited is local maximum. It is usually used in reconstructing electromagnetic showers and jets from tau-lepton decays. The second is a topological algorithm, which starts with a seed cell and adds to the cluster the neighbour of a cell already in the cluster iteratively, if the significance of the to-be-added cell's energy over the expected background noise is above a threshold. The algorithm is good at suppressing noise in the clusters containing large number cells and is usually used for jet and E_T^{miss} reconstruction.

Sometimes, the cluster of towers is used instead of the cluster of cells. To build a tower, the calorimeter is divided into a grid, in the $\eta - \phi$ space, of $N_\phi \times N_\eta$ elements of some predefined size. These elements are the tower. Inside each tower, the energy of all cells in all longitudinal layers is summed into the tower energy.

3.4.4 Electron and photon

In the central region of the calorimeter system ($|\eta| < 2.47$), the electron and photon reconstruction starts with the matching of a track to an EM cluster whose total transverse energy above 2.5 GeV, by requiring the η distance between them to be less than 0.05 and ϕ distance to be less than 0.1. In case when multiple tracks are matched to the same EM cluster, tracks with hits in the silicon detector are preferred and the closest in terms of ΔR is chosen.

If the EM cluster does not match any track in the ID, it is reconstructed as an unconverted photon, otherwise as an electron. If the cluster matches a pair of tracks originating from a reconstructed conversion vertex, the two tracks will be

reconstructed as two electrons that are from an converted photon and the cluster will be reconstructed as a converted photon.

The energy of the EM clusters that are reconstructed as electrons or photons is computed by a weighted sum of four different contributions in the EM calorimeter: the energy deposit in the material in front of the EM calorimeter, the energy deposit in the cluster, the external energy deposit outside the cluster (lateral leakage) and the energy deposit beyond the EM calorimeter (longitudinal leakage).

In the forward region where only EM calorimeters can be used, the reconstruction of electrons relies on the shower shape of the EM cascade.

After the reconstruction of electrons and photons, there are further cut-based selection to identify true electrons or photons and reject fake ones, e.g. jets. For electrons, there are three sets of customised cuts (*loose*, *medium*, and *tight*) with increasing jet rejection power. These cuts make use of shower shapes in the EM calorimeter as well as TRT radiation to identify electron. For photons, only two sets of cuts (*loose* and *tight*) are developed. The details of the variables used in the cuts and the performance of the identification can be found in [25, 26]. Recently, a likelihood method with several optional working point based on multivariable analysis for electron identification is available, which will be used in later analysis [27].

3.4.5 Muon

Muons are reconstructed using the information from the muon spectrometer, inner detector and calorimeter. There are several types of muons available:

- Stand-Alone (SA) muons: the muon track is only reconstructed in the MS. The parameters of the muon track at the IP are determined by extrapolating back to the point of closest approach to the beam line, taking into account the estimated energy loss in the calorimeters. SA muons are mainly used to extend the acceptance to the range $2.5 < |\eta| < 2.7$, which is not covered by the ID,

- ComBined (CB) muon: the muon is reconstructed through a combination of the tracks in the ID and MS, which is the main type of reconstructed muons and has the highest purity,

- Segment-Tagged (ST) muons: the muon is reconstructed by a combination of the track in the ID and at least one track segment in the MDT or CSC. ST muons can be used to increase acceptance in cases in which the muon crossed

only one layer of the MS, either because of its low p_T or because it falls in regions with reduced MS acceptance,

- Calorimeter-Tagged (CaloTag) muons: the muon is reconstructed by a combination of the track in the ID and an energy deposit in the calorimeter. This type has the lowest purity but recovers acceptance in the uninstrumented regions of the MS.

There are two sets of algorithms available for the muon reconstruction: *Muid* algorithm and *Staco* algorithm. The corresponding reconstructed muons are called separately *Muid* muon and *Staco* muon. A third algorithm exists which combine the above two and produce a unified muon chain. It is planned to be used as the single algorithm in the LHC "RUN II" (see Ch. B).

After reconstruction, there are a set of hit requirements on the ID track of the muon from the muon combined performance (MCP) group to further consolidate the muon quality: the sum of pixel hits and dead pixel hits sensors crossed by the track should be greater than zero; the sum of SCT hits and dead SCT sensors crossed by the track must be greater than four; the number of missing hits in a crossed sensor which is not dead ("holes") must be less than three; and a successful TRT extension must be found if the track is within the acceptance of the TRT.

3.4.6 Jets and b-tagging

An energetic quark or gluon will radiate quarks or gluons during its flight until their energy is below the threshold where hadron formation starts. All the hadrons formed are called in simplicity a jet. In ATLAS, jets are very common, either from hard process or the initial/final state radiation.

The ATLAS jet reconstruction takes as input the calorimeter towers and clusters to build calorimeter jet. Since the default cell energy is reconstructed at EM scale, this kind of jet is called EM jet. There is a local cluster weighting (LCW) method which calibrates clusters based on cluster properties related to shower development. Jet reconstructed using these calibrated clusters is called LCW jet. The reconstructed ID tracks can also be used in jet reconstruction, since charged hadrons in a jet will leave track information in the ID, to build track jet. Besides, particles from event generator (which will be introduced in detail in later chapter) can be used to build the so-called truth jet.

There are two sets of algorithms for the jet reconstruction: the FastJet [28] and the infrared safe seedless cone algorithm SISCone [29]. The default jet reconstruction algorithm is the anti-k_T algorithm with full four-momentum recombination ("E-scheme") and two jet sizes, $R = 0.4$ and $R = 0.6$, are available.

After the jet is reconstructed (from calorimeter), sequences of corrections are applied like: pile-up correction which subtracts the pile-up energy from the jet energy; event vertex correction which corrects the change of direction induced by the fact that primary event vertex could be displaced from the origin of the reference frame; etc.

Jets initiated from bottom quarks are called b-jets. The identification of these quarks are called b-tagging, which is very important for the selection of top quark sample or the rejection of top quark background.

The b-jets possess several characteristic properties that can be utilised to separate them from jets coming from lighter quarks. The most important one is the relatively long lifetime of b-hadrons of about 1.5 ps, which corresponds to a flight length of a few mm before their decay that can be measured.

There are various b-tagging methods in ATLAS which can be divided into two classes: the spatial taggers which comprise methods that use lifetime information and the soft taggers that are based on the reconstruction of the lepton which is from the semi-leptonic decay of the b-hadron. All taggers are based on multivariable technique and will give a weight (likelihood ratio) to each jet. If the weight is above a certain cut value, the jet is tagged as b-jet. This cut value, also called working point of the tagger, determines the efficiency of b-tagging as well as the rejection rate of light-jet, which are p_T and η dependent.

Besides, b-tagging is limited in the range $|\eta| < 2.4$ due to the need of tracking information in the ID.

3.4.7 Missing transverse energy

The missing energy in ATLAS comes from those non-interacting particles like neutrinos or potential BSM particles. They can be indirectly detected and measured by implying conservation in the transverse plane (there is no momentum or energy information in the beam direction due to the fact that only a fraction of the protons' energy, which is unknown and described by the PDFs, is available in the collision):

$$\sum_{\text{interacting}} \vec{p}_T + \sum_{\text{jets}} \vec{p}_T + \sum_{\text{non-interacting}} \vec{p}_T = 0. \tag{3.5}$$

The default E_T^{miss} is calorimeter-based $E_T^{\text{miss, calo}}$. It is calculated by taking the negative transverse energy sum of all the observable objects, which are the electron, muon, tau, jet, photon and those cells not belonging to these objects (denoted as $E_T^{\text{miss, CellOutEflow}}$). There is a derived quantity called $E_T^{\text{miss, calo, STVF}}$, where a pile-up suppression technique is applied for rejecting pile-up jets and for correcting pile-up energy in $E_T^{\text{miss, CellOutEflow}}$.

There is also a track-based missing transverse energy called $E_T^{\text{miss, track}}$, which is calculated mainly from tracks of hard objects (electrons, muons and jets) and other soft tracks. Thus it has little dependence on the pile-up. Based on $E_T^{\text{miss, track}}$, there is another derived quantity called $E_T^{\text{miss, track, jetCorr}}$, of which the calculation of jet energy is changed from track momentum to calorimeter energy.

For each type of missing transverse energy, there exists a projected version of it, onto the axis of the nearest hard object. Take the $E_T^{\text{miss, calo}}$ as example, it is defined as:

$$E_{T,\text{Rel}}^{\text{miss, calo}} = \begin{cases} E_T^{\text{miss, calo}}, \text{ if } \Delta\phi_{\text{closest}} \geqslant \pi/2 \\ E_T^{\text{miss, calo}} \times \sin\Delta\phi_{\text{closest}}, \text{ if } \Delta\phi_{\text{closest}} < \pi/2\,. \end{cases} \tag{3.6}$$

Chapter 4

Data and Monte Carlo samples

The best way to examine the correctness of a theory is to setup an experiment to check the prediction of the theory. In high energy physics (HEP), the experimental results are collected as data samples, which are always in huge amount and in which the new physics might hide. As for the prediction of the theory, due to the random nature of elementary particles, we use Monte Carlo (MC) simulation to model these random processes. By comparing the data sample with MC sample, we could valid/reject a theory or set confidence interval for the parameters in a theory.

In this chapter, the data and Monte Carlo samples, relevant to the physics analyses to be presented in the thesis, will be introduced.

4.1 Luminosity

Before introducing the data and MC samples, we will give the definition of the quantify "luminosity", which is to quantify the rate at which the physics processes occur.

Take as an example the elastically scattered particles on a fixed target, the number of total scattered particles is denoted as N. Usually we are interested in the number of scattered particles in a given solid angle per unit of time, which can be written as a product of two terms:

$$\frac{d^2N}{d\Omega dt} = L\frac{d\sigma}{d\Omega}\,. \tag{4.1}$$

In the above formula, $\frac{d\sigma}{d\Omega}$ is the probability density distribution for the particle to be scattered in a given direction, which describes the physics. The proportionality

factor L is called the **luminosity**. It corresponds to the flux of projectiles, thus describes the experimental conditions. If we integrate the above formula over the full solid angle, we have:

$$R = \frac{dN}{dt} = L\sigma\,, \tag{4.2}$$

where R is the scattering rate of a given process which is described by its cross section σ. In this case, the luminosity could be interpreted as the event rate of a physics process.

If we further integrate the above formula over a given time, we get the total number of scattered particles or interactions in an experiment. This is called integrated luminosity $\mathcal{L} = \int L dt$, which is measured in inverse femtobarns (fb^{-1}), a unit roughly equivalent to 100 trillion collisions. It is this quantity we usually use to quantify the size of a data sample.

As an example, the cumulative luminosity as a function of time, delivered to ATLAS for the first three-year run is shown in Fig. 4.1.

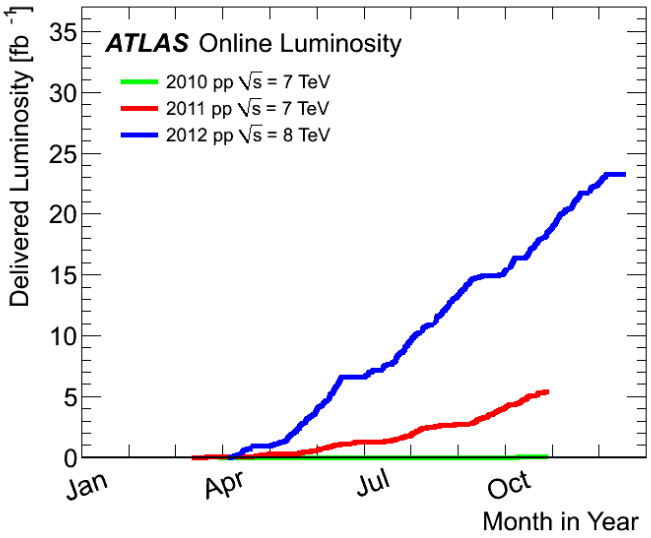


Figure 4.1: Cumulative luminosity versus day delivered to ATLAS during stable beams and pp collisions in year 2010 (green), 2011 (red) and 2012 (blue).

4.2 Data sample

The data sample used in the following analyses was collected in years 2011 at a centre of mass energy of $\sqrt{s} = 7\,\mathrm{TeV}$ and an integrated luminosity of $4.5\,\mathrm{fb}^{-1}$ and

Table 4.1: Period dependent trigger setup in the year 2011 used in the analysis.

Period	ee channel	$\mu\mu$ channel	eμ channel
B - I	EF_e20_medium	EF_mu18_MG	EF_e20_medium || EF_mu18_MG
J	EF_e20_medium	EF_mu18_MG_medium	EF_e20_medium || EF_mu18_MG_medium
K	EF_e22_medium	EF_mu18_MG_medium	EF_e22_medium || EF_mu18_MG_medium
L - M	EF_e22vh_medium1	EF_mu18_MG_medium	EF_e22vh_medium1 || EF_mu18_MG_medium

in year 2012 at a centre of mass energy of $\sqrt{s} = 8\,\text{TeV}$ and an integrated luminosity of $20\,\text{fb}^{-1}$.

As we mentioned before, when introducing the trigger and DAQ system of the ATLAS, events are collected by triggers and stored according to the type of triggers they satisfy. In the case of $H \to WW^* \to \ell\nu\ell\nu$, there are two high-$p_T$ leptons (electron or muon) in the final state. Thus we use single and dilepton triggers.

In 2011, only single lepton triggers were used, which are summarised in Table 4.1. The data sample is separated into different periods due to changing detector conditions. So the triggers are changing accordingly. In the table, "EF" means event filter, which is the third level trigger. "e" and "mu" represent the electron trigger and muon trigger. The numbers after "e" and "mu" mean the nominal p_T threshold values in GeV. The suffix "medium" indicates the tightness in the electron identification criteria, "vh" means the trigger has both η dependent p_T threshold and a hadronic leakage cut at the LV1 trigger.

In 2012, both single and dilepton triggers were used, which are summarised in Table 4.2. In the table, "i" after "vh" means that the isolation criteria (which will be introduced later) are used. The suffixes "medium1", "loose1", and "tight" mean the tightness of the identification. The number "2" before "e" means that it is a di-electron trigger.

Table 4.2: Trigger setup for 2012 run.

ee channel	EF_e24vhi_medium1 || EF_e60_medium1||EF_2e12Tvh_loose1 ||EF_2e12Tvh_loose1_L2StarB
$\mu\mu$ channel	EF_mu24i_tight || EF_mu36_tight||EF_mu18_tight_mu8_EFFS
$e\mu$ & μe channels	EF_e24vhi_medium1 || EF_e60_medium1 || EF_mu24i_tight || EF_mu36_tight ||EF_e12Tvh_medium1_mu8

The trigger efficiency is measured in order to estimate the true number of events. This is done by the combined performance group using tag and probe method with Z data, and has been found to be $\sim 90\%$ for electrons and $\sim 90\%$ ($\sim 70\%$) for muons in the end-cap (barrel).

Besides, we have used supporting triggers to measure the lepton fake factors, which will be described when we introduce the estimation of W+jets background. For the muon fake factor measurement, the EF_mu6 is used for $10 < p_T < 15\,\text{GeV}$ and EF_mu15 for $p_T > 15\,\text{GeV}$. For the electron fake factor measurement, the medium1 triggers together with the "EtCut" are used.

4.3 Monte Carlo sample

The Monte Carlo sample is in fact pseudo data sample generated by computer using programs modelling the pp collision and the detector response. By repeating the programs, in principle, we can have infinite number of pseudo experimental data. By generating MC samples of various processes, the composition of a real data sample can be estimated.

4.3.1 Sample production

The production of a MC sample involves roughly two steps: the generation of truth level events and the simulation of detector response.

The truth level event generation centres on the desired hard process and starts from the proton-proton collision to the stage where only stable particles exist in the final products.

The program responsible for the generation is called event generator. The workflow of a typical event generator is shown in Fig. 4.2. It starts with the generation of the hard process, which in this case is a W+jets process, as shown in Fig. 4.2(a). In this step, the PDFs introduced before are used by the program to extract momentum fractions for the two incoming partons. Then initial state radiation (ISR) or final state radiation (FSR) is added, which is shown in green and blue in Fig. 4.2(b). These radiations are usually softer than those from the hard process. Then the underlying events from multiple parton interactions (MPI) are added, which are the long black helices from the protons and their attached ISR or FSR in Fig. 4.2(c). These processes are described by non-perturbative QCD and thus relying on phenomenological modelling of the program. At last, all the

partons produced in the former steps form hadrons, which is called hadronisation, and subsequently the unstable hadrons decay.

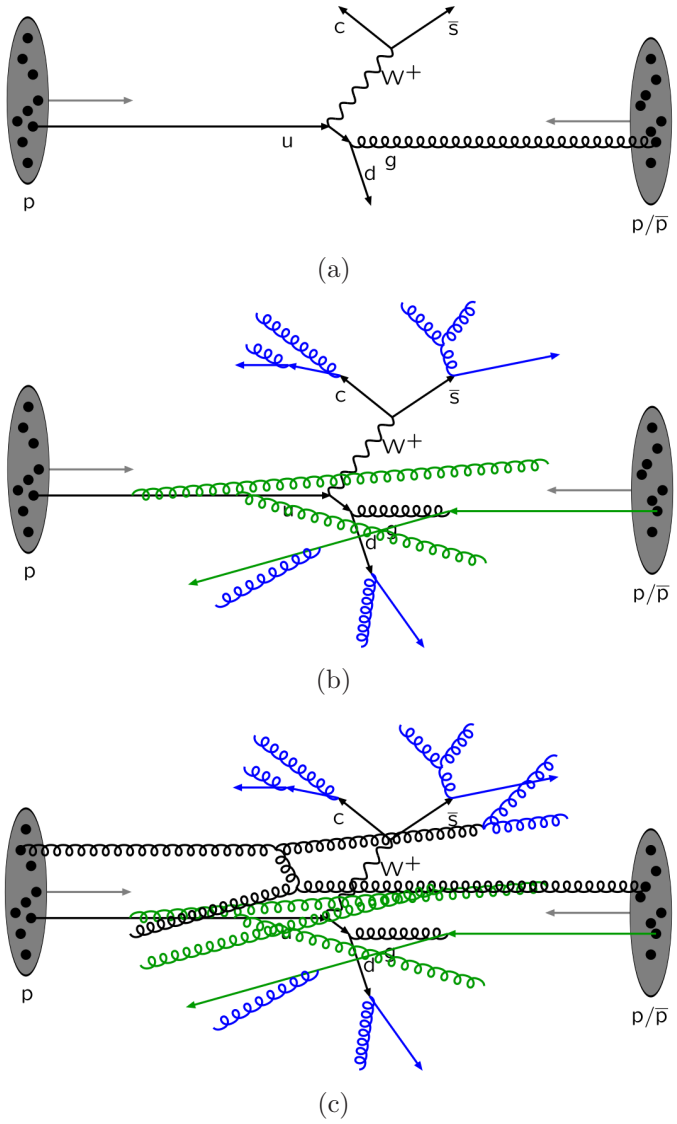


Figure 4.2: The workflow of a typical event generator: (a) hard process; (b) ISR/FSR; (c) MPI.

Sometimes the whole procedures are integrated in one event generator, while sometimes there are dedicated generators called Matrix Element (ME) generator responsible for the generation of a hard process and interfaced to Parton Shower (PS) programs finishing the rest parts of the job. In the latter case, depending on the perturbative order of the hard process calculated, there are LO or NLO ME generator. Some example NLO generators are: MC@NLO [30], HERWIG [31] and POWHEG [32–34].

Truth level events, which consist of information on the four momenta of the final state particles, are fed into the detector simulation programs, which are based on Geant4 [35, 36], to simulate the passage of those particles through the detector.

The simulation results are the hits in active volumes of the detector, which need further processing by the so-called digitisation program to be converted as Raw Data Objects, the input to the reconstruction softwares. During the digitisation, pile-up events, such as minimum bias, cavern background, beam halo and beam gas interactions, can be added.

The final step of the MC sample generation is thus reconstruction of the event that mentioned in Ch. 3.

4.3.2 List of MC samples

The MC samples to be used in the following analyses in the $H \to WW^* \to \ell\nu\ell\nu$ channel are introduced in this section. Different MC samples corresponding to various processes are produced to analyse the composition of the real data sample. Those processes we interested in are called signal, in this case the Higgs production and decay, while the others called background. In the following, leptonic decays of W/Z bosons are always assumed and cross sections include the branching ratios and are summed over lepton flavours. Besides, Higgs signal samples are all produced at $m_H = 125\,\text{GeV}$.

The signal MC samples include [37, 38]:

- The dominant ggF process is modelled by POWHEG [39]+PYTHIA8.
- The second-dominant VBF process is also modelled by POWHEG+PYTHIA8.
- The Higgs-Strahlung WH/ZH process is modelled by PYTHIA.
- Other small production processes, like ttH and bbH, are neglected because of their tiny contribution.
- The various samples of spin-2 Higgs boson are generated by MadGraph5_aMC@NLO [40], which includes higher-order tree-level QCD calculations.
- The ggF signal $gg \to H \to WW$ used for off-shell analysis is generated by MCFM in LO precision and showered by PYTHIA8. Besides, the same generator is used for the production of the $gg \to WW$ sample and another sample, $gg \to (H^*) \to WW$, which includes both the ggF signal and the $gg \to WW$ background as well as their interference.

- The VBF signal used for off-shell analysis is generated by MadGraph+PYTHIA8, and so are the related background and signal plus background plus interference samples.

The background samples include [37]:

- The continuum WW production ($q\bar{q}/g \to WW$) is modelled to NLO precision by POWHEG+PYTHIA6. While the small contribution from gluon-gluon initiated quark box diagram ($gg \to WW$) is modelled via gg2VV+HERWIG, which is in LO precision. In high jet multiplicity analyses, WW+jets is modelled in Sherpa (LO) due to the second jets coming from parton showers in POWHEG sample is poorly modelled.

- The $t\bar{t}$ process is modelled by POWHEG+PYTHIA6 (NLO). The relevant single-top production channels are also included: s-channel and Wt are modelled also by POWHEG+Pythia6 (NLO), while t-channel is modelled by AcerMC+PYTHIA6.

- The inclusive QCD Z/γ^* process (also referred to as Drell-Yan process) is modelled by ALPGEN+HERWIG, while the EW inclusive Z/γ^* process is modelled by Sherpa.

- The inclusive W+jets process is modelled by ALPGEN+HERWIG. But this sample is only used for fake factor systematic evaluation and subtraction in QCD background estimation.

- For $WZ^{(*)}$ process, POWHEG+PYTHIA8 is used and for $W\gamma^{(*)}$ process, Sherpa is used. Besides, the interference between these two processes is included. $Z^{(*)}Z^{(*)} \to l\nu l\nu$ process is modelled in POWHEG+PYTHIA8. EW WZ/ZZ+ 2jets processes is generated in Sherpa. For $W\gamma$ process, ALPGEN is used. For $Z\gamma$ process, Sherpa is used. All these processes are named as Non-WW diboson process.

- Double Parton Interaction of W is modelled by PYTHIA8.

In the spin analysis, the same background samples as those in the coupling analysis are used, while additional signal samples are generated:

Different PDF sets are used together with different event generators: the CT10 PDF set is used for the POWHEG, Sherpa and gg2VV samples; the CTEQ6L1 is

used for the ALPGEN, AcerMC, PYTHIA8 and several Sherpa in 7 TeV samples, but with the ALPGEN Z/γ^* sample reweighted to the MRSTCal.

All the above information on MC samples and their corresponding generators and cross sections are summarised in Table 4.3.

Table 4.3: MC generators used to model the signal and background processes in the coupling analysis, and the corresponding cross sections (given for $m_H = 125\,\text{GeV}$ in the case of the signal processes). Leptonic decays of W/Z bosons are always assumed, and the quoted cross sections include the branching ratios and are summed over lepton flavours.

Process	Generator	$\sigma \cdot \text{Br}(8\text{TeV})$ (pb)	$\sigma \cdot \text{Br}(7\text{TeV})$ (*pb*)
ggF $H \to WW$	POWHEG [41]+PYTHIA8 [42]	0.435	0.341
VBF $H \to WW$	POWHEG [43]+PYTHIA8	$36 \cdot 10^{-3}$	$28 \cdot 10^{-3}$
WH/ZH $H \to WW$	PYTHIA8 (PYTHIA6)	$25 \cdot 10^{-3}$	$21 \cdot 10^{-3}$
Spin-2 signal samples:			
$\kappa_g = \kappa_q = 1$ $H \to WW$	MadGraph5_aMC@NLO+PYTHIA6	-	
$\kappa_g = 1, \kappa_q = 0$ $H \to WW$	MadGraph5_aMC@NLO+PYTHIA6	-	
$\kappa_g = 0.5, \kappa_q = 1$ $H \to WW$	MadGraph5_aMC@NLO+PYTHIA6	-	
Off-shell signal/background samples:			
$gg \to H \to WW$	MCFM+PYTHIA8	0.07652	
$gg \to WW$	MCFM+PYTHIA8	0.08110	
$gg \to (H^*) \to WW$	MCFM+PYTHIA8	0.15058	
$VBF \to H \to WW$	MadGraph [44]+PYTHIA	0.02160	
$VBF \to WW$	MadGraph+PYTHIA	0.01809	
$VBF \to (H^*) \to WW$	MadGraph+PYTHIA	0.03770	
$VBF \to (H^*) \to WW$ $(10\times\mu_{\text{off-shell}})$	MadGraph+PYTHIA	0.04047	
$q\bar{q}/g \to WW$	POWHEG+PYTHIA6	5.68	4.68
$gg \to WW$	GG2WW [45]+HERWIG	0.20	0.14
QCD $WW + 2$ jets	Sherpa	0.568	-
EW $WW + 2$ jets	Sherpa	0.039	0.027
$t\bar{t}$ dileptonic	POWHEG [46]+PYTHIA6	26.6	18.6
tW/tb leptonic	POWHEG [46]+PYTHIA6	4.17	3.15
tqb leptonic	AcerMC [47]+PYTHIA6	28.4	20.7
inclusive W	ALPGEN [48]+HERWIG	$37 \cdot 10^3$	$31 \cdot 10^3$
inclusive $Z/\gamma^\star$ ($m_{ll} \geq 10\,\text{GeV}$)	ALPGEN [48]+HERWIG	$16.5 \cdot 10^3$	$14.9 \cdot 10^3$
EW $Z/\gamma^\star$	Sherpa	5.36 (inc. t-ch)	2.26
$W(Z/\gamma^*)$	POWHEG+PYTHIA8	12.7	10.8
$W(Z/\gamma^*)$ ($m_{(Z/\gamma^*)} < 7\,\text{GeV}$)	Sherpa	12.2	10.6
$Z^{(*)}Z^{(*)} \to 4l(2l2\nu)$	POWHEG+PYTHIA8	0.73(0.50)	0.64(0.42)
EW $WZ + 2$ jets	Sherpa	$13 \cdot 10^{-3}$	$8.5 \cdot 10^{-3}$
EW $ZZ + 2$ jets ($4l, ll\nu\nu$)	Sherpa	$73 \cdot 10^{-5}(12 \cdot 10^{-4})$	$53 \cdot 10^{-5}(8.8 \cdot 10^{-4})$
$W\gamma$	ALPGEN [48]+HERWIG	369	313
$Z\gamma(p_T^\gamma > 7\,\text{GeV})$	Sherpa	163	-
DPI WW	PYTHIA8	0.440	-

4.3.3 Reweighting

The produced MC samples may not well agree with real data in some aspects, like the pile-up condition, the integrated luminosity, trigger efficiencies, etc. In these occasions, reweighting methods are usually used to tune the MC samples to be the data sample.

Pile-up reweighting

In most 8 TeV MC samples, the pile-up modelling is defined before data-taking, which means that the pile-up condition in the MC sample is different from that of real data sample. This is shown in Fig. 4.3.

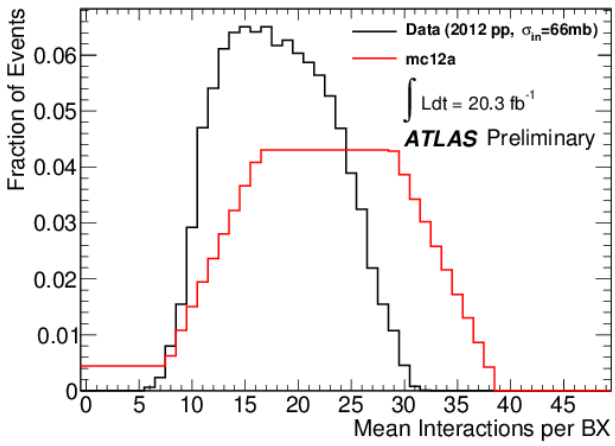


Figure 4.3: Average interaction per bunch-crossing distribution defined in mc12a compared with in real data.

To obtain a realistic pile-up treatment, we reweight MC sample to data by a 0.9 μ rescaling, where μ is the average number of interactions per bunch crossing.

In the generation of 7 TeV MC sample, the pile-up condition is already tuned to agree with that of data, thus no μ rescaling is needed.

Luminosity reweighting

The number of events of a MC sample is usually at least one order of magnitude larger than that of the expected number of events for the same process in the data sample, which is to reduce statistical uncertainty from the MC sample. To get a correct prediction on the expected number of events of that process to be compared with the data, the MC sample needs to be scaled to the same integrated luminosity as the data by applying an overall scale (weight) to all the MC events according to the formula:

$$w = \frac{\mathcal{L} \cdot \sigma}{N}, \tag{4.3}$$

where $\mathcal{L}$ is the integrated luminosity of the data and σ is the cross section of the MC process.

Trigger scale factor

The trigger simulation, which is a part of the detector simulation, could bring bias to the MC sample, if its efficiency is different from that of the real trigger. In this case, a per-lepton trigger scale factor, defined as the ratio of the trigger efficiency in MC over that in the real data and usually p_T/η dependent, can be used to correct the difference if there is only one lepton in the event providing the trigger.

In the $H \to WW^* \to \ell\nu\ell\nu$ channel, there could be two high p_T leptons, either of which could fire the trigger. Thus the per-event trigger scale factor is a bit complicated and calculated using these per-lepton trigger scale factors according to the formula:

$$\text{per-event SF} = \frac{\epsilon^{\text{event}}_{\text{data}}}{\epsilon^{\text{event}}_{\text{MC}}} = \frac{1 - (1 - \epsilon^{\text{lead}}_{\text{MC}} \times \text{SF}^{\text{lead}}) \times (1 - \epsilon^{\text{sub}}_{\text{MC}} \times \text{SF}^{\text{sub}})}{1 - 1 - \epsilon^{\text{lead}}_{\text{MC}} \times 1 - \epsilon^{\text{sub}}_{\text{MC}}}\,, \tag{4.4}$$

where $\epsilon^{\text{lead}}_{\text{MC}}$ and $\epsilon^{\text{sub}}_{\text{MC}}$ are the per-lepton trigger efficiencies for the lepton with leading p_T and the lepton with sub-leading p_T in an event, and SF^{lead} and SF^{sub} are the per-lepton scale factors for the leading and sub-leading lepton, respectively. In the case of dilepton trigger, the ϵ^{event} in the above formula needs to be modified according to the following formula:

$$\epsilon^{\text{event}} = \epsilon_{\text{single lep}} + \epsilon_{\text{dilep}} - \epsilon_{\text{single lep}} \cdot \epsilon_{\text{dilep}}\,. \tag{4.5}$$

There are other reweighting factors to be applied in practice, e.g. scale factors for the correction of lepton reconstruction and identification efficiencies in MC, which will not be explained unless encountered in the following.

Chapter 5

Observation of the Higgs boson

In 2012, the Higgs boson was discovered with contributions from three analysis channels: ZZ, $\gamma\gamma$, and WW [1]. At that time, it was the combined results of the three that allowed the Higgs discovery (the observed significance $> 5\,\sigma$). While in each single channel, either only the evidence of Higgs boson was observed (the observed significance $> 3\,\sigma$) in ZZ^* and $\gamma\gamma$, or not even enough for providing evidence in WW^* (the observed significance was $2.8\,\sigma$) [1].

Since then, the integrated luminosity increased, as the LHC keeps delivering collision data, for $\sqrt{s} = 8\,\text{TeV}$ from 5.8 fb^{-1} to 20 fb^{-1} which reduced the statistical uncertainty of the analysis. Also the $H \to WW^* \to \ell\nu\ell\nu$ (HWW) analysis undergoes a series of developments, in two time scales from July 2012 to Moriond 2013 and from Moriond 2013 to now, in almost every aspect of the original analysis. We have improved the sensitivity and reduced the systematic uncertainties. All these efforts have helped to increase the expected significance from $2.8\,\sigma$ to $5.8\,\sigma$ and the observed one from $2.8\,\sigma$ to $6.1\,\sigma$, which means that the Higgs boson is now observed independently in the $H \to WW^* \to \ell\nu\ell\nu$ channel alone [49].

5.1 Analysis overview

The aim of the analysis is to isolate the low mass Higgs bosons decaying into W boson pair and subsequently two leptons and two neutrinos from a sea of various SM processes. This is realised by imposing, on the data sample, cuts which are carefully designed so that a maximum sensitivity of the $H \to WW^* \to \ell\nu\ell\nu$ process can be reached.

The Higgs boson production mode considered in the analysis is the SM Higgs

boson produced at $m_H = 125\,\text{GeV}$, via ggF (red in Fig. 3.7), VBF (blue in Fig. 3.7), and VH (purple in Fig. 3.7), each with decreasing contribution.

In the decay chain of $H \to WW^* \to \ell\nu\ell\nu$ ($l = e, \mu$), the branching ratio of WW for a SM Higgs boson at $m_H = 125\,\text{GeV}$ is 22%, the largest branching ratio except for $b\bar{b}$ at this mass and the dilepton decay mode of the W pair occurs with a probability of 10.5%. These considerable probabilities make the WW channel powerful for the observation of the Higgs boson.

The decay products of the W boson pair, two high-p_T leptons and neutrinos, can be used to tag the candidate events. Thus the analysis starts from the data with two isolated (the definition of isolation will be given later) charged opposite-sign leptons in the final state. Based on the possible flavour combination of the two leptons, the analysis is divided into sub-channels of **same flavour channel** ($ee, \mu\mu$) and **different flavour channel** ($e\mu, \mu e$, where the first lepton is the one with leading p_T). Besides, the kinematic properties (spin correlation) of the two leptons can be used to identify signal events. As for the two neutrinos in the final state, they give a non-zero E_T^{miss} which can be used to reject backgrounds.

Jets will appear in the final state of the signal processes, either from ISR or FSR. It is found that the compositions of data samples with different jet multiplicity are quite different, which can be seen in Fig. 5.1 [37]. The yellow band represents the statistical uncertainty only. That's why the KS probability is quite low (the KS probability is the results of the Kolmogorov-Smirnov test of the equality of two distributions [50]). Thus the analysis is further divided into sub-channels according to the jet multiplicity of the data sample: $\mathbf{n_j = 0}$ **channel** and $\mathbf{n_j = 1}$ **channel**, which are dominated by ggF signals, and **VBF** $\mathbf{n_j \geq 2}$ **channel**, which is dominated by VBF signals. In the VBF channel, the topology of the two leading jets (also called tagging jet) could be used to tag the signal. Besides, there is a **ggF** $\mathbf{n_j \geq 2}$ **channel**, which is orthogonal to the VBF channel and optimised for the search of the ggF signals, to gain additional sensitivity.

The main backgrounds can be classified into five categories top ($t\bar{t}$ and Wt), WW, non-WW dibosons ($W\gamma$, $W\gamma^{(*)}$, WZ, ZZ, and $Z\gamma$), misidentification (W+jets and QCD), and Drell-Yan (Z/γ^*+jets). The top backgrounds have a similar final state signature as the signal if they decay to WW pair. However, they can be suppressed by tagging the b quark, which is produced from top decay together with W boson. The continuum WW background is quite similar to the signal process in terms of final state products and is called irreducible background. The non-WW dibosons (also called "VV" have two or more than two real leptons in their final

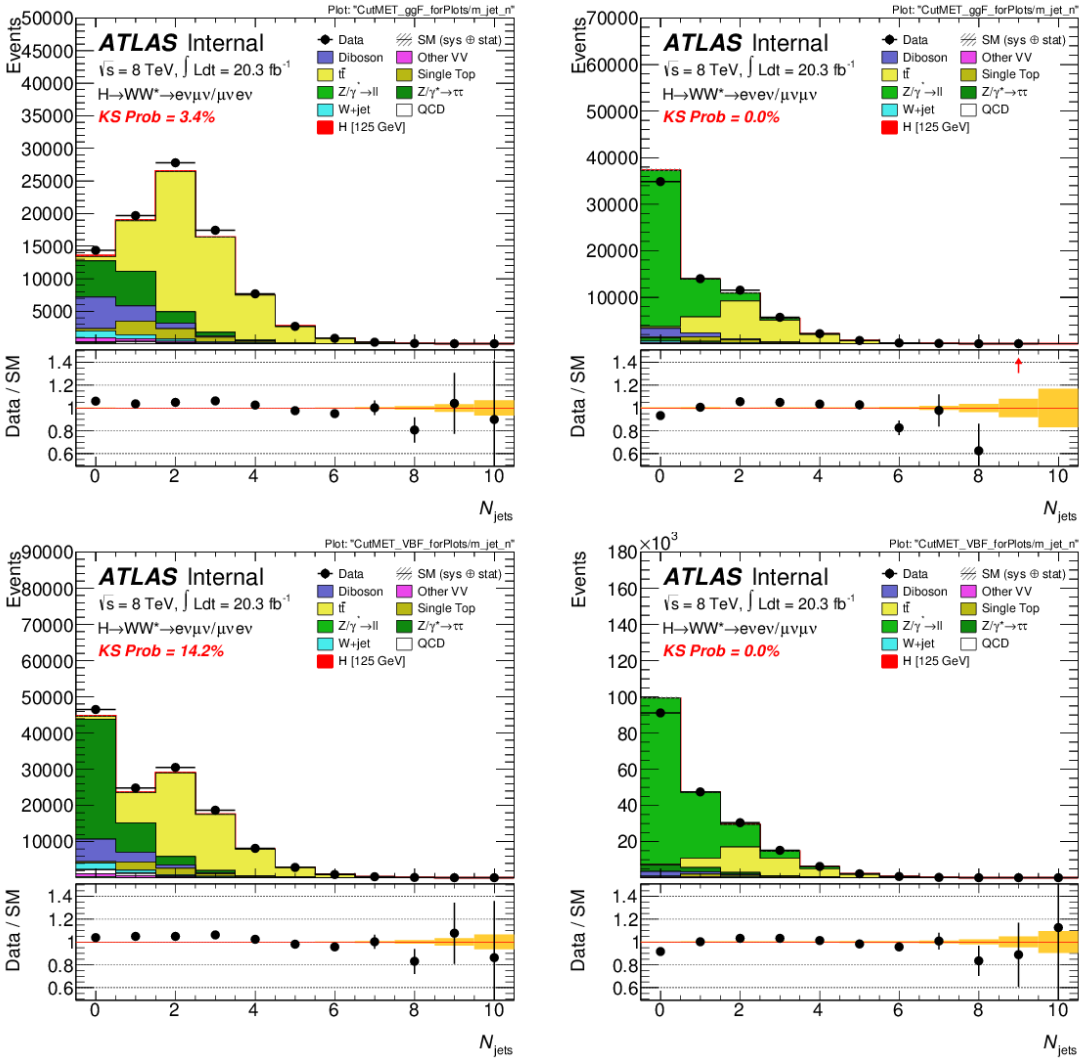


Figure 5.1: The jet multiplicity distribution after the dilepton preselection and $E_{\mathrm{T}}^{\mathrm{miss}}$ selection for the different flavour (left) and same flavour (right) lepton pairs. The top pair of plots shows the composition in the ggF-enriched analysis and the bottom pair shows the equivalent for the VBF-enriched analysis.

Table 5.1: Main Backgrounds to the $H \to WW^* \to \ell\nu\ell\nu$ analysis in the final state. Irreducible backgrounds have the same final state; other backgrounds are shown with the features that lead to this final state. Quarks from the first or second generation are denoted as q, and j represents a jet of any flavour.

Name	Process	Feature(s)
WW	WW	Irreducible
Top quarks		
$t\bar{t}$	$t\bar{t}\to Wb\,W\bar{b}$	Unidentified b-quarks
single top (st)	tW	Unidentified b-quark
	$t\bar{b}$, $tq\bar{b}$	q or b misidentified as ℓ; unidentified b-quarks
Misidentified leptons (Misid)		
W+jets	W+jets	jet misidentified as ℓ
QCD	Multijet	jets misidentified as $\ell\ell$; misidentified neutrinos
Other dibosons		
VV	$W\gamma$	γ misidentified as e
	$W\gamma^*, WZ, ZZ\to\ell\,\ell$	Unidentified lepton(s)
	$ZZ\to\ell\ell\,\nu\nu$	Irreducible
	$Z\gamma$	γ misidentified as e; unidentified lepton
Drell-Yan (DY)		
$ee/\mu\mu$	$Z/\gamma^* \to ee, \mu\mu$	Misidentified neutrinos
$\tau\tau$	$Z/\gamma^* \to \tau\tau \to \ell\nu\nu\,\ell\nu\nu$	Irreducible

state. When the additional leptons are unidentified or the photon is misidentified as electron to make up the missing one, these non-WW dibosons come into play. Besides, if the ZZ process has one Z to lepton pair and the other to neutrino pair, it constitutes another irreducible background. The W+jets and QCD, which are denoted as "misidentification", pass the selection due to the jets falsely identified as leptons. Finally, the $ee/\mu\mu$ decay channel of the Drell-Yan process ($Z/\gamma^* \to ee/\mu\mu$), constitutes an important background to the same flavour analysis, because of E_T^{miss} from the mis-measurement of objects in the event. As for the $Z/\gamma^* \to \tau\tau$ decay process, which is mainly relevant to different flavour analysis, has genuine missing energy and two leptons in its final state due to leptonic τ decays and is also irreducible. All these backgrounds are summarised in Table 5.1 [49]. In the following analyses, they will be normalised using data-driven methods as much as possible.

After the various event selections to be presented in the following, a profile

likelihood fit will be performed to extract the signal from the backgrounds and measure the signal strength μ, which is defined as the signal event rate relative to the SM prediction. Based on the measurement of μ, we can draw conclusions on whether the Higgs signal is observed and if the observation is in agreement with the SM prediction.

In the following, the analysis will be introduced and especially those parts I involved in. Only the 8 TeV analysis is described in great detail, while the 7 TeV analysis will be mentioned simply when necessary.

5.2 Objects and observables

The event selection in an analysis is usually based on cuts on some variables having discriminating power between signal and background. These variables are introduced in this section.

The physical objects, like leptons, jets and E_T^{miss}, are the candidate variables to be used. They are reconstructed by the reconstruction softwares and selected by some predefined identification criteria mentioned before. But there are always more than one choice of the objects provided, thus the optimal choice should be made in specific analysis, Besides, there may be additional analysis-dependent requirements on these objects to fit the practical needs. These choices and requirements are described below. Except for these basic objects, high-level quantities constructed from them are also useful in the analysis, whose definition will be given in the following.

5.2.1 Leptons

Tighter lepton selection than the standard ATLAS lepton identification criteria is applied in the HWW analysis, which is mainly for the purpose of rejecting the W+jets and QCD backgrounds, with the corresponding tradeoff for the signal efficiency. The additional selection refers to the transverse and longitudinal impact parameter as well as calorimeter- and track-based isolation, which will be explained in the following. The criteria are dependent on the lepton E_T, because the rejection of the W+jets background is the dominant consideration, and the contribution from this background drops off sharply with increasing lepton p_T [37].

Electron

For electrons with $E_T < 25\,\text{GeV}$, the "Very Tight Likelihood" (VTLH) identification is used because it provides better rejection of background from non-prompt electrons than the tightest cut-based identification, but retains similar efficiency. For electrons with $E_T > 25\,\text{GeV}$, the "Medium++" cut-based identification is used, with two requirements modified to improve the rejection of electrons from photon conversion: for all electrons, the conversion flag is required to be false and the associated track must have a hit in the innermost pixel layer if expected. In the default Medium++, these requirements are only applied to central ($|\eta| < 2.37$) electrons.

The requirements on impact parameters are universal in all E_T regions: $d_0/\sigma_{d_0} < 3.0$, where d_0 is the transverse impact parameter and σ_{d_0} its measured uncertainty; $z_0 \sin\theta < 0.4\,\text{mm}$, where z_0 is the longitudinal impact parameter and θ the polar angle.

Isolation cuts are used to distinguish prompt leptons from very short-lived particle decays, like W and Z boson, against jets misidentified as leptons. The calorimeter isolation energy is computed by summing the transverse energy of clusters with positive energy in a cone of size $\Delta R = 0.3$ around the electron and then removing the energy of those EM cells in a rectangle area around the electron to avoid including the energy of the electron itself. The track isolation energy, which has a better performance than the calorimeter one, is computed by summing the transverse momentum of the tracks in a cone of size $\Delta R = 0.4(0.3)$ around the track of the electron. Details of the isolation definition can be found in Ref. [27]. Both isolation energies will be normalised to the transverse energy of the electron.

The impact parameter cuts and isolation cuts are summarised in Table 5.2 [37].

Table 5.2: Electron selection as a function of E_T. "CBL" refers to the conversion flag and b-layer hit requirements extended to all η (within the electron acceptance coverage).

E_T (GeV)	Electron ID	Calo. isolation topoEtConeCor	Track isolation Ptcone	Impact parameters
10-15	Very Tight LH	$(\text{iso}(0.3))/E_T < 0.20$	$(\text{iso}(0.4))/E_T < 0.06$	$d_0/\sigma_{d_0} < 3.0$, $z_0 \sin\theta < 0.4\,\text{mm}$
15-20		$(\text{iso}(0.3))/E_T < 0.24$	$(\text{iso}(0.3))/E_T < 0.08$	
20-25		$(\text{iso}(0.3))/E_T < 0.28$	$(\text{iso}(0.3))/E_T < 0.10$	
> 25	Medium++ with "CBL"			

Besides, electron kinematic cuts are applied:

- The electrons are required to be within the geometrical acceptance of the detector, which is $|\eta| < 2.47$ except for the transition region ($1.37 < |\eta| < 1.52$) between the barrel and end-caps of the calorimeter.
- The electrons are required to have a large transverse momentum $p_T > 10\,\text{GeV}$.

To account for the difference in the selection efficiency of the above cuts between data and MC samples, scale factors should be applied. The reconstruction and identification scale factors as well as their uncertainties are provided by performance groups, while the rest are measured with the Z tag-and-probe method.

Muon

The *Staco* muon is used and ID track hit requirements from the MCP group are applied. Similar impact parameter cuts and isolation cuts are imposed, which are summarised in Table 5.3 [37]. Their scale factors and uncertainties are also either from performance groups or measured by the Z tag-and-probe method. Besides, muon kinematic cuts are also applied: $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$.

Table 5.3: Muon selection as a function of E_T.

p_T (GeV)	Calo. isolation EtConeCor	Track isolation Ptcone	Impact parameters
10-15	$(\text{iso}(0.3))/E_T < 0.06$	$(\text{iso}(0.4))/E_T < 0.06$	$d_0/\sigma_{d_0} < 3.0$, $z_0 \sin\theta < 1.0$ mm
15-20	$(\text{iso}(0.3))/E_T < 0.12$	$(\text{iso}(0.3))/E_T < 0.08$	
20-25	$(\text{iso}(0.3))/E_T < 0.18$	$(\text{iso}(0.3))/E_T < 0.12$	
> 25	$(\text{iso}(0.3))/E_T < 0.30$		

Two leptons are considered overlapped if the angular distance in the $\eta - \phi$ plane is too small: $\Delta R < 0.1$. There is an overlap removal mechanism to protect against the same object being reconstructed as two electrons or one electron and muon. In the former case, the electron with higher p_T is kept and in the latter case, the muon is always kept. If the overlap between electron and muon is too serious as $\Delta R < 0.05$, the whole event will be dropped.

5.2.2 Jets and b-jets

The LCW jets reconstructed by Anti-k_T4 of cone size $\Delta R = 0.4$ are used to be consistent with the jets used in the calculation of E_T^{miss}. After optimisation, the following requirements are applied [37]:

- LOOSER jet cleaning, which is to reject bad jets not associated to real energy deposits in the calorimeters,
- $|\eta| < 4.5$,
- $p_T > 25(30)\,\text{GeV}$ for $|\eta| < (\geq)2.4$,
- $|\text{JVF}| > 0.5$ for $p_T < 50\,\text{GeV}$ and $|\eta| < 2.4$, where JVF represents for Jet-Vertex-Fraction and is a discriminating variable between pile-up jet and hard scatter jet.

Similar as the lepton overlap removal, there is a jet-lepton removal rule with the threshold on ΔR loosened to 0.3. If it is an electron getting close to the jet to that extent, the jet is removed. Otherwise, if is is a muon, the muon is removed.

For the b-jet tagging, the MV1 tagger working at 85% b-jet efficiency is used. In most cases, jets with $p_T > 20\,\text{GeV}$ are used as input to the tagger (the special case is for the JVSP method, which will be introduced later, of the top estimation in the $n_j = 0$ channel where the nominal jet p_T threshold is used).

5.2.3 Missing transverse energy

There are many flavours of E_T^{miss} available as mentioned in Sec. 3.4.7. The choice, after optimisation study, depends on the lepton flavour combination and jet multiplicity [37]:

- For the different flavour channel, the E_T^{miss} with the best resolution was chosen - $E_T^{\text{miss, track, jetCorr}}$,
- For the same flavour channel, the E_T^{miss} with the best rejection power of Drell-Yan was chosen - $E_{T,Rel}^{\text{miss, track}}$,
- For the VBF channel, no E_T^{miss} cut gives the best performance for different flavour channel and for the same flavour channel, the best performance was provided by a combination of $E_T^{\text{miss, calo}}$ and $E_T^{\text{miss, track, jetCorr}}$.

And the above results are summarised in Table 5.4.

Table 5.4: Different E_T^{miss} flavours used in the $H \to WW^* \to \ell\nu\ell\nu$ analysis.

Jet bin	E_T^{miss} flavour	Description
0j $e\mu + \mu e$	$E_T^{\text{miss, track}}$	track-based E_T^{miss} with calorimeter jets
0j $ee + \mu\mu$	$E_{T,\text{Rel}}^{\text{miss, track}}$, $E_{T,\text{Rel}}^{\text{miss}}$	projections of the track- and calo-based E_T^{miss}
1j $e\mu + \mu e$	$E_T^{\text{miss, track, jetCorr}}$	track-based E_T^{miss} with calorimeter jets
1j $ee + \mu\mu$	$E_{T,\text{Rel}}^{\text{miss, track}}$, $E_{T,\text{Rel}}^{\text{miss}}$	projections of the track- and calo-based E_T^{miss}
2j/VBF $e\mu + \mu e$	-	-
2j/VBF $ee + \mu\mu$	$E_T^{\text{miss, calo}}$, $E_T^{\text{miss, track, jetCorr}}$	calo-based and track-based E_T^{miss} with calorimeter jets

5.2.4 Common observables

Common observables, constructed from the kinematic variables of the above objects and used in the whole $H \to WW^* \to \ell\nu\ell\nu$ analysis, are defined in this section.

- Lepton related variables:
  - $m_{\ell\ell}$: the dilepton invariant mass,
  - $p_{T,\ell\ell}$: the dilepton transverse momentum,
  - $\Delta\phi_{\ell\ell}$: the dilepton opening angle in the transverse plane,
  - $\Delta\eta_{\ell\ell}$: the absolute difference of the pseudorapidity of the two leptons,
  - $m_{\tau\tau}$: the $\tau\tau$ invariant mass, calculated by assuming the two leptons are all from τ decays based on the so-called collinear approximation [51],
  - $M_{T,\text{max}}^W$: the max possible W boson transverse mass, calculated by $M_T^W = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos(\varphi^\ell - \varphi^{E_{\text{T}}^{\text{miss}}}))}$.
- Missing transverse energy related variables:
  - $E_T^{\text{miss, calo}}$ ($E_{T,\text{Rel}}^{\text{miss,calo}}$): calorimeter-based missing transverse energy (projected onto the nearest hard object),
  - $E_T^{\text{miss, calo, STVF}}$ ($E_{T,\text{Rel}}^{\text{miss, calo, STVF}}$): pile-up suppressed calorimeter-based missing transverse energy (projected onto the nearest hard object),
  - $E_T^{\text{miss, track}}$ ($E_{T,\text{Rel}}^{\text{miss, track}}$): track-based missing transverse energy (projected onto the nearest hard object),
  - $E_T^{\text{miss, track, jetCorr}}$ ($E_{T,\text{Rel}}^{\text{miss, track, jetCorr}}$): jet energy corrected calorimeter-based missing transverse energy (projected onto the nearest hard object).

- Jet related variables:
  - n_j: the number of jets,
  - $N_{b\text{-jet}}$: the number of b-jets,
  - ΔY_{jj}: (VBF channel) the rapidity gap between the two leading jets,
  - m_{jj}: (VBF channel) the dijet invariant mass,
  - CJV: (VBF channel) Central Jet Veto, which is a boolean flag and takes the value of *true* if there is a jet with $p_T > 20\,\text{GeV}$ lies between the two leading jets in η direction. More specifically, the centrality C_{j_3} of the third jet with respect to the two leading jets is defined as:

$$C_{j_3} = 2 \cdot \left| \frac{\eta_{j3} - \bar{\eta}}{\eta_{j_0} - \eta_{j_1}} \right|, \tag{5.1}$$

$$\tag{5.2}$$

    where η_{j_0}, η_{j_1}, and η_{j_3} are pseudorapidities of the three jets and the $\bar{\eta}$ is the average of the former two. The CJV is *true* when $C_{j_3} \leq 1$.

- Variables constructed from leptons and E_T^{miss} system:
  - $\Delta\phi_{\ell\ell,E_T^{\text{miss}}}$: the azimuthal angle between the dilepton system and the E_T^{miss},
  - f_{recoil}: ($n_j = 0$ channel) p_T of the recoil system (soft jets) with respect to the leptons and E_T^{miss} system, normalised to the p_T of the latter.

- Variable constructed from leptons and jets system:
  - OLV: (VBF channel) Outside Lepton Veto, which is a boolean flag and takes the value of *true* (thus the event should be vetoed) if there is a lepton lies outside the rapidity gap of the two leading jets. More specifically, the centrality of the two leptons are defined by:

$$\mathrm{C}_{\ell 1} = 2 \cdot \left| \frac{\eta_{l_0} - \bar{\eta}}{\eta_{j_0} - \eta_{j_1}} \right| \tag{5.3}$$

$$\mathrm{C}_{\ell 2} = 2 \cdot \left| \frac{\eta_{l_1} - \bar{\eta}}{\eta_{j_0} - \eta_{j_1}} \right|, \tag{5.4}$$

    where η_{ℓ_0} and η_{ℓ_1} are the pseudorapidities of the two leptons. The OLV is *true* when either $\mathrm{C}_{\ell 1}$ or $\mathrm{C}_{\ell 2}$ is larger than one,

– $\sum C_\ell$: (VBF channel) an extension of the OLV that quantifies the overall positions of the two leptons with respect to the two leading jets in the η-plane:

$$\sum C_\ell = \mathrm{C}_{\ell 1} + \mathrm{C}_{\ell 2}. \tag{5.5}$$

- Variables constructed from the leptons, jets, and E_T^{miss} system:
  - $f_{\mathrm{recoil}}^{\mathrm{extended}}$: ($n_j = 1$ channel) strength of the recoil system (soft jets) with respect to the leptons, jets, and E_T^{miss} system, normalised to the transverse momentum of the latter,
  - p_T^{tot}: the transverse momentum of all the identified hard objects,
  - m_T: the transverse mass of the whole system, defined as:

$$m_T = \sqrt{(E_{\mathrm{T}}^{\ell\ell} + E_T^{\mathrm{miss}})^2 - |\boldsymbol{p}_T^{\ell\ell} + \boldsymbol{E}_T^{\mathrm{miss}}|^2}, \tag{5.6}$$

    where $E_{\mathrm{T}}^{\ell\ell} = \sqrt{|\boldsymbol{p}_T^{\ell\ell}|^2 + m_{\ell\ell}^2}$,
  - $\sum_{\ell,j} M_{\ell j}$: (VBF channel) the sum of the invariant masses of all four possible lepton-jet pairs (from two leptons and two jets).

5.3 Event selection

With all the variables to be used defined in the previous section, channel-dependent event selections will be described in detail in the following sections.

5.3.1 Common preselection

As mentioned at the beginning of this chapter, the $H \to WW^* \to \ell\nu\ell\nu$ analysis is further divided into sub-analyses by the lepton flavour combination and jet multiplicity. These sub-analyses channels are optimised with different set of cuts on different variables. However, before the analysis branches, there is common preselection [37]:

- Primary vertex (PV) selection, the PV should be consistent with the beam spot position and has at least three associated tracks with $p_T > 400\,\mathrm{MeV}$ and have the largest $\sum(p_T)^2$, which is summed over all tracks associated with the PV,

- Trigger match, at least one of the selected lepton should match with the object that fires the trigger,
- Exactly two opposite sign leptons,
- $p_T^{\text{lead}} > 22\,\text{GeV}$, $p_T^{\text{sublead}} > 10\,\text{GeV}$, to reject mainly the W+jets and QCD background. Besides, due to the low mass Higgs ($\sim 125\,\text{GeV}$) considered in the analysis, among its two intermediate decay products of W bosons, one is off-shell (which explains the notation W^*) giving a softer lepton than that from the on-shell W boson. The sub-leading lepton p_T cut used to be set at 15 GeV due to the large backgrounds in the [10, 15] GeV (low-p_T) bin. However, after a dedicated study of these low-p_T events which I involved in, it is found that these events could give significant contributions to the nominal analysis in particular with the use of the likelihood electron identification and the improved W+jets and QCD background treatment at low-p_T.
- $m_{\ell\ell} > 10$ (12) GeV for different (same) flavour lepton channel, to remove mainly the low mass Drell-Yan background,
- $|m_{\ell\ell} - m_Z| > 15\,\text{GeV}$, to remove the Z Drell-Yan background,
- Channel-dependent E_T^{miss} cuts: to suppress backgrounds out of mis-measured E_T^{miss}, like the Drell-Yan and QCD (since they do not have high p_T neutrinos in their final state) as summarised in Table 5.5.

Table 5.5: E_T^{miss} flavours and their cut thresholds (in GeV) applied in the $H \to WW^* \to \ell\nu\ell\nu$ analysis.

Jet multiplicity	$n_j = 0$		$n_j = 1$		$n_j \geq 2$ VBF BDT(cut-based)		$n_j \geq 2$ ggF
Lepton flavours	$ee/\mu\mu$	$e\mu/\mu e$	$ee/\mu\mu$	$e\mu/\mu e$	$ee/\mu\mu$	$e\mu/\mu e$	$e\mu/\mu e$
$E_{T,\text{Rel}}^{\text{miss, track}}$	40	–	35	–	–	–	–
$E_T^{\text{miss, track, jetCorr}}$	–	20	–	20	50(40)	–	20
$E_T^{\text{miss, calo}}$	–	–	–	–	55(45)	–	–
$E_{T,\text{Rel}}^{\text{miss, calo}}$	40	–	40	–	–	–	–

After the common preselection, events are categorised into different channels according to jet multiplicity, lepton flavuor combination, and signal process. First of all, events are classified into three jet bins: $n_j = 0$, $n_j = 1$, $n_j \geq 2$. Especially, the $n_j \geq 2$ channel is separated into two channels: one for the ggF signal and the other for the VBF signal study. Then each jet bin is divided into same flavour $e\mu$

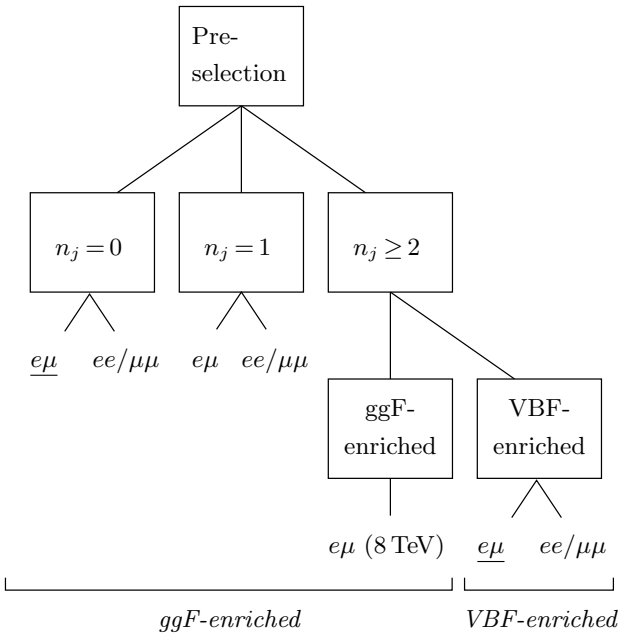


Figure 5.2: Event categorisation based on jet multiplicity (n_j) and lepton-flavour samples ($e\mu$ and $ee/\mu\mu$). The most sensitive signal region for ggF production is $n_j = 0$ in $e\mu$, while for VBF production it is $n_j \geq 2$ in $e\mu$. These two samples are underlined.

and different flavour $ee/\mu\mu$. Except that in the $n_j \geq 2$ channel for ggF study, only $e\mu$ flavour is analysed. All these descriptions are summarised in the block diagram of Fig. 5.2 [49].

5.3.2 ggF $n_j = 0$ channel

Around 75% of the ggF signal, which is the main production mode of the Higgs boson, falls into the event category without any jet due to the high-p_T threshold of the jet selection and the soft nature of its radiation jet. Thus this is the channel that has the most sensitivity to the Higgs boson.

The following cuts are applied:

- $n_j = 0$: channel-defining cut, which suppressed the top quark background by two order of magnitude,
- $\Delta\phi_{\ell\ell,E_T^{\text{miss}}} > \pi/2$: to further remove events with mis-measured E_T^{miss} since if there is no mis-measurement, the dilepton system and the E_T^{miss} tend to be back-to-back to get balance,

- $p_{T,\ell\ell} > 30\,\text{GeV}$: to further reduce Drell-Yan (by an order of magnitude) as shown in Fig. 5.3(a) [49] since the two leptons from Drell-Yan decay tend to be oppositely aligned thus have small $p_{T,\ell\ell}$,

- $m_{\ell\ell} < 55\,\text{GeV}$: motivated by the spin correlation of the two leptons which is explained in Fig. 5.4 [49], to reject continuum WW background, but also effective to other backgrounds, as shown in Fig. 5.3(b) [49],

- $E_{T,\text{Rel}}^{\text{miss, track}} > 40\,\text{GeV}$: only applied to $ee/\mu\mu$ channels: to further reject Drell-Yan events,

- $\Delta\phi_{\ell\ell} < 1.8$: motivated again by the spin correlation of the two leptons to further suppress Drell-Yan process, which can be seen in Fig. 5.3(c) [49]. Besides, the $m_{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ cuts are called the Higgs topological selection,

- $f_{\text{recoil}} < 0.1$: only applied to $ee/\mu\mu$ channels: to further remove Drell-Yan background as shown in Fig. 5.3(d) [49], since the soft recoil energy in these events tend to be large to balance the transverse momentum, while for the signal and other backgrounds with true missing energy part of the recoil energy is taken away by the non-interacting particles resulting in a smaller recoil energy.

The cutflow corresponds to the above cuts is shown in Table 5.6 [49]. After all the above selection, there are 209 (73) expected signal events over 2350 (1096) expected background events left in the $e\mu/\mu e$ ($ee/\mu\mu$) channels. There is an additional row in the cutflow table, which is a simple m_T cut. From it, a rough signal-over-background (SOB) ratio $\sim 10\%$ can be calculated to have a feeling about the final sensitivity achieved by the whole set of cuts. Besides, the transverse mass m_T, whose distribution will be used as input to the final fit, is shown in Fig. 5.5(a)(b) [49] for the $e\mu/\mu e$ and $ee/\mu\mu$ channel. In the case of $e\mu/\mu e$, the fit is performed in six signal regions by dividing firstly the original signal region into three regions, in terms of p_T^{sublead}, at [10, 15], [15, 20], and [20, ∞] GeV. Then each of the three is further divided into two regions in terms of $m_{\ell\ell}$ at [10, 30] and [30, 55] GeV.

5.3.3 ggF $n_j = 1$ channel

Compared to the $n_j = 0$ event channel, less ggF but more VBF signal falls into this channel. And the overall amount of signal events in this channel is significantly

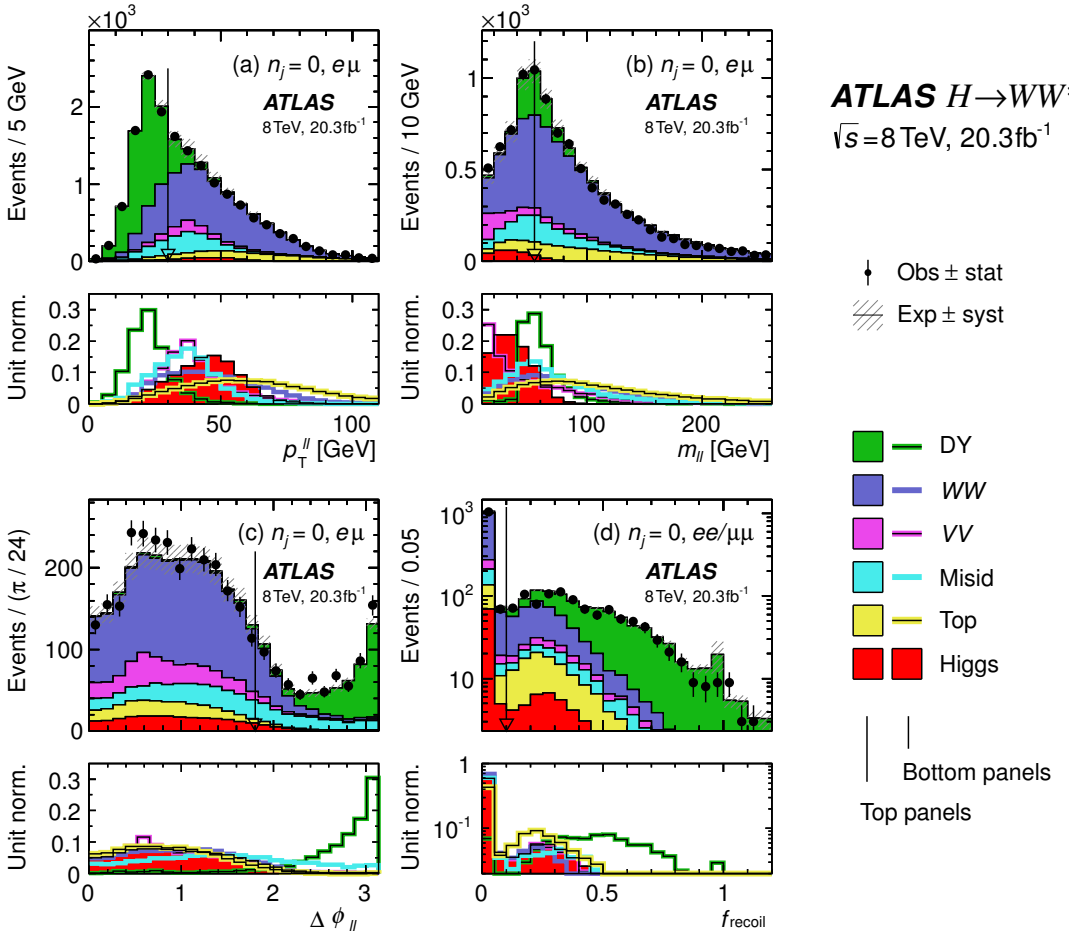


Figure 5.3: Distributions of (a) $p_{T,\ell\ell}$, (b) $m_{\ell\ell}$, (c) $\Delta\phi_{\ell\ell}$, and (d) f_{recoil}, for the $n_j = 0$ category. The plot in (a) is made after requiring all selections up to $p_{T,\ell\ell}$, (b) up to $m_{\ell\ell}$, (c) up to $\Delta\phi_{\ell\ell}$, and (d) up to f_{recoil}. For each variable, the top panel compares the observed and the cumulative expected distributions; the bottom panel shows the overlay of the distributions of the individual expected contributions, normalised to unit area, to emphasise shape differences.

less than that in $n_j = 0$ channel. However, after the following optimised event selection, almost the same level of SOB ratio can be achieved:

- $n_j = 1$: channel-defining cut,
- $n_{b\text{-jet}} = 0$: b-jet veto cut, also called "bveto", to reject the top quark background.
- $M_{T,\text{max}}^{W} > 50\,\text{GeV}$, only applied to $e\mu/\mu e$ channels: to control the Drell-Yan and multijet background effectively as shown in Fig. 5.6(a) [49],

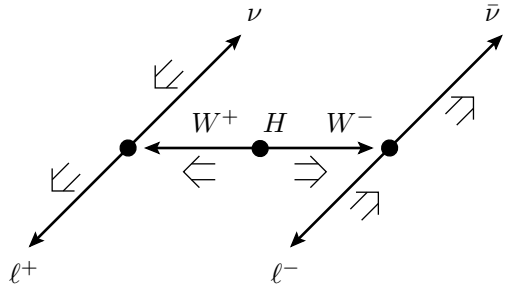


Figure 5.4: Illustration of the $H \to WW^* \to \ell\nu\ell\nu$ decay. The small arrows indicate the particles' directions of motion and the large double arrows indicate their spin projections. The spin-0 Higgs boson decays to W bosons with opposite spins, and the spin-1 W bosons decay into leptons with aligned spins. The H and W boson decays are shown in their rest frames. Because of the $V-A$ decay of the W bosons, the charged leptons have a small opening angle in the laboratory frame. This feature is also present when one W boson is off-shell which is the case of low mass Higgs decay.

Table 5.6: Cutflow for event selection of the $n_j = 0$ category in the 8 TeV data analysis. The selection is presented separately for the $e\mu/\mu e$ and $ee/\mu\mu$ channels. The summary columns give the observed yields (N_{obs}), the expected background yields (N_{bkg}), their ratios, and the expected signal yields (N_{sig}). Background normalisations are applied. The N_{sig} values are given for $m_H = 125\,\text{GeV}$ and are subdivided into the N_{ggF} and N_{VBF} contributions. The composition columns give the contributions to N_{bkg}. Entries are shown as 0.0 (−) if they are less than 0.1 (0.01) events. The entries are rounded to a precision commensurate with the statistical uncertainties. Energy-related quantities are in GeV.

Selection	Summary					Composition of N_{bkg}							
	N_{obs}/N_{bkg}	N_{obs}	N_{bkg}	N_{sig}		N_{WW}	N_{top}		N_{misid}		N_{VV}	N_{DY}	
				N_{ggF}	N_{VBF}		$N_{t\bar{t}}$	N_{st}	N_{W+jets}	N_{QCD}		$N_{ee/\mu\mu}$	$N_{\tau\tau}$
$e\mu/\mu e$ sample	1.01±0.01	16423	16330	290	12.1	7110	820	407	1330	237	739	115	5570
$\Delta\phi_{\ell\ell,E_T^{\text{miss}}} > \pi/2$	1.00±0.01	16339	16270	290	12.1	7110	812	405	1330	230	736	114	5530
$p_{T,\ell\ell} > 30$	1.00±0.01	9339	9280	256	10.3	5690	730	363	1054	28	571	60	783
$m_{\ell\ell} < 55$	1.11±0.02	3411	3060	224	6.3	1670	141	79	427	12	353	27	350
$\Delta\phi_{\ell\ell} < 1.8$	1.12±0.02	2642	2350	203	5.9	1500	132	75	278	9.2	324	19	12
$\frac{3}{4}m_H < m_T < m_H$	1.20±0.04	1129	940	131	2.2	660	40	21	133	0.8	78	4.3	2.3
$ee/\mu\mu$ sample	1.04±0.01	38040	36520	163	7.2	3260	418	211	504	29	358	31060	685
$\Delta\phi_{\ell\ell,E_T^{\text{miss}}} > \pi/2$	1.05±0.01	35445	33890	163	7.1	3250	416	211	493	26	355	28520	622
$p_{T,\ell\ell} > 30$	1.06±0.01	11660	11040	154	6.8	3010	394	201	396	2.6	309	6700	21
$m_{\ell\ell} < 55$	1.01±0.01	6786	6710	142	5.0	1260	109	64	251	2.0	179	4840	8.7
$E_{T,\text{Rel}}^{\text{miss, track}} > 40$	1.02±0.02	2197	2160	117	4.3	1097	99	59	133	0.5	106	660	0.3
$\Delta\phi_{\ell\ell} < 1.8$	1.01±0.02	2127	2100	113	4.2	1068	96	57	122	0.5	104	649	0.3
$f_{\text{recoil}} < 0.1$	1.01±0.03	1108	1096	72	2.7	786	41	31	79	0.0	69	91	0.1
$\frac{3}{4}m_H < m_T < m_H$	0.99±0.05	510	517	57	1.3	349	11	8	53	-	31	64	0.1

- $m_{\tau\tau} < m_Z - 25\,\text{GeV}$, only applied to $e\mu/\mu e$ channels: to reduce significantly the remaining Drell-Yan events as shown in Fig. 5.6(b) [49]. This is also called $Z\tau\tau$ veto,

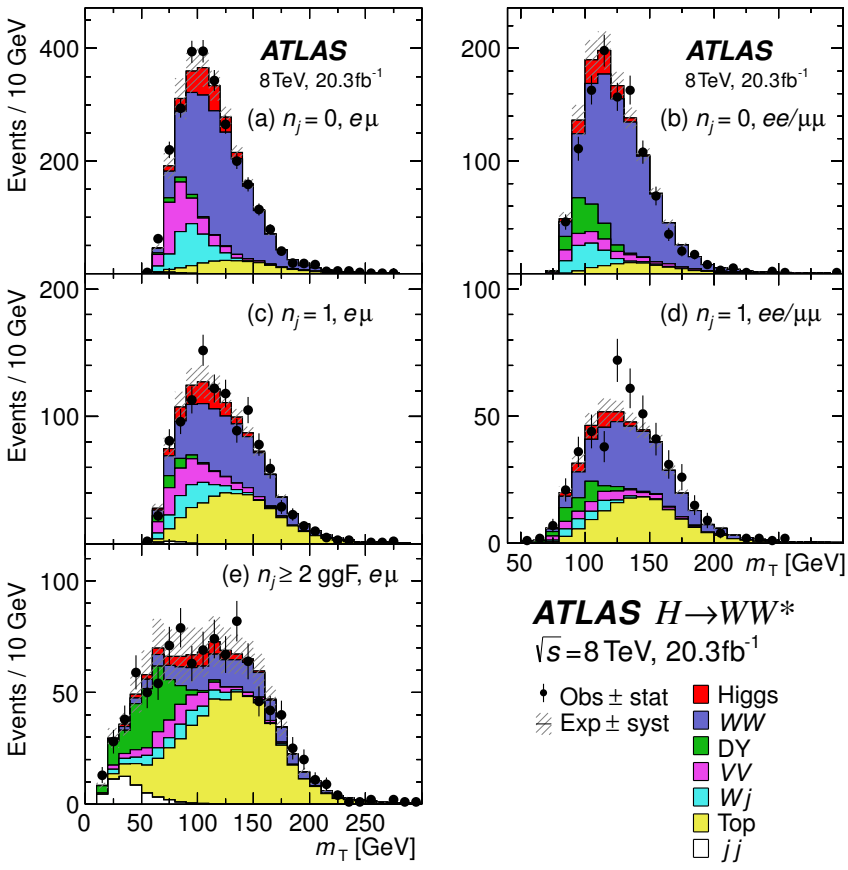


Figure 5.5: Distributions of the transverse mass m_T after all cuts up to the m_T for the ggF $n_j = 0$, $n_j = 1$ and $n_j \geq 2$ channels in the 8 TeV data analysis. jj means QCD and Wj means W+jets.

- $m_{\ell\ell} < 55\,\text{GeV}$: topological cut which is shown in Fig. 5.6(c) [49],
- $E_{T,\text{Rel}}^{\text{miss, track}} > 35\,\text{GeV}$: only applied to $ee/\mu\mu$ channels: similar as in the $n_j = 0$ channel,
- $\Delta\phi_{\ell\ell} < 1.8$: topological cut which is shown in Fig. 5.6(d) [49],
- $f_{\text{recoil}}^{\text{extended}} < 0.1$: only applied to $ee/\mu\mu$ channels: similar as in the $n_j = 0$ channel Fig. 5.3(d) [49].

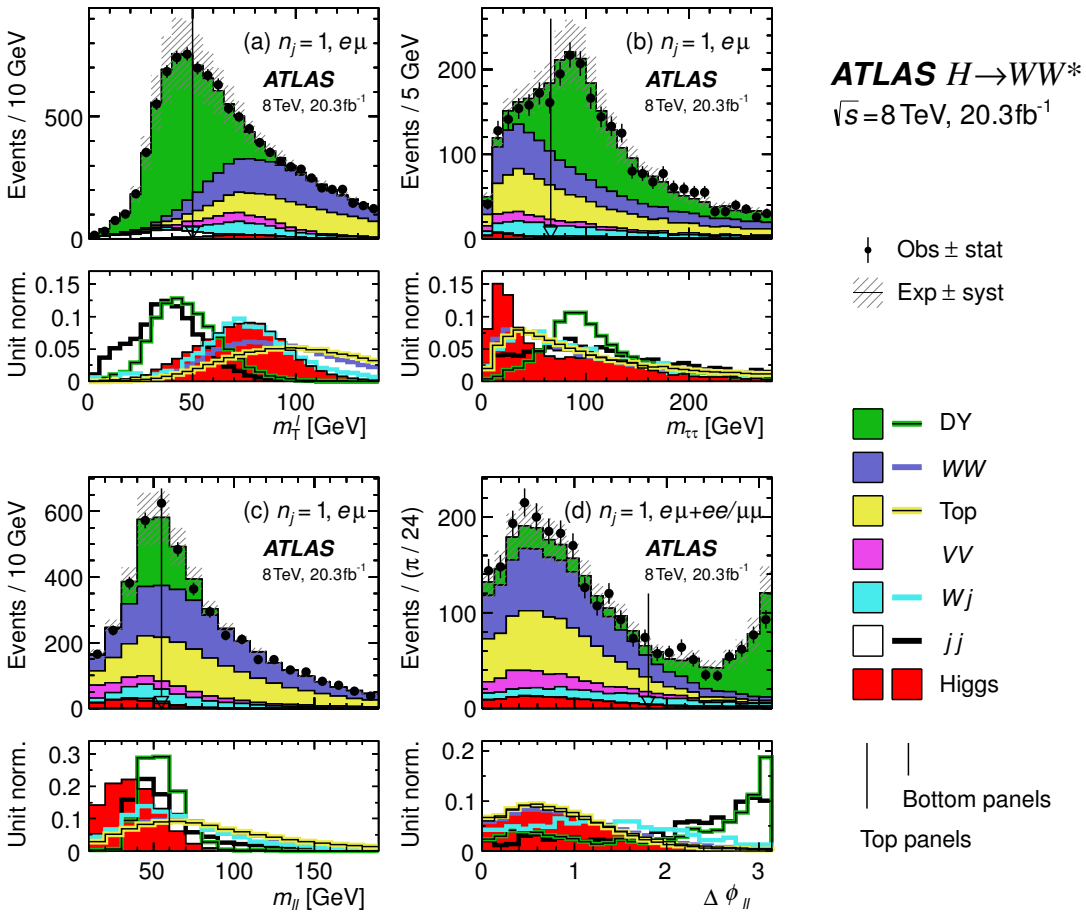


Figure 5.6: Distributions of (a) $M_{T,\text{max}}^W$, (b) $m_{\tau\tau}$, (c) $m_{\ell\ell}$, and (d) $\Delta\phi_{\ell\ell}$, for the $n_j = 1$ category. The plot in (a) is made after requiring all selections up to $m_{\tau\tau}$, (b) up to $m_{T,\text{max}}^W$, (c) up to $m_{\ell\ell}$, and (d) up to $\Delta\phi_{\ell\ell}$.

The cutflow corresponds to the above cuts is shown in Table 5.7 [49]. After all the above selection, there are 87 (24) expected signal events over 1030 (404) expected background events left in the $e\mu/\mu e$ ($ee/\mu\mu$) channels. From the additional

Table 5.7: Cutflow for event selection of the $n_j = 1$ category in the 8 TeV data analysis. (see Table 5.6 for presentation details).

Selection	Summary					Composition of N_{bkg}							
	N_{obs}/N_{bkg}	N_{obs}	N_{bkg}	N_{sig}		N_{WW}	N_{top}		N_{misid}		N_{VV}	N_{DY}	
				N_{ggF}	N_{VBF}		$N_{t\bar{t}}$	N_{st}	N_{W+jets}	N_{QCD}		$N_{ee/\mu\mu}$	$N_{\tau\tau}$
$e\mu/\mu e$ sample	1.00±0.01	20607	20700	131	32	2750	8410	2310	663	334	496	66	5660
$n_{bj} = 0$	1.01±0.01	10859	10790	114	26	2410	1610	554	535	268	423	56	4940
$M^W_{T,\max} > 50$	1.01±0.01	7368	7280	103	23	2260	1540	530	477	62	366	43	1990
$m_{\tau\tau} < m_Z - 25$	1.02±0.02	4574	4490	96	20	1670	1106	390	311	32	275	21	692
$m_{\ell\ell} < 55$	1.05±0.02	1656	1570	84	15	486	297	111	129	19	139	6.4	383
$\Delta\phi_{\ell\ell} < 1.8$	1.10±0.03	1129	1030	74	13	418	269	102	88	6.1	119	5.0	22
$\frac{3}{4}m_H < m_T < m_H$	1.21±0.06	407	335	42	6.6	143	76	30	40	0.5	42	1.1	2
$ee/\mu\mu$ sample	1.05±0.01	15344	14640	61	15	1111	3770	999	178	13	192	8100	280
$n_{bj} = 0$	1.08±0.02	9897	9140	53	12.1	972	725	245	137	10	163	6640	241
$m_{\ell\ell} < 55$	1.16±0.02	5127	4410	48	9.4	351	226	85	73	7.8	79	3420	168
$E^{\text{miss, track}}_{T,\text{Rel}} > 35$	1.14±0.04	960	842	36	6.9	292	193	73	38	0.2	49	194	2
$\Delta\phi_{\ell\ell} < 1.8$	1.14±0.04	889	783	32	6.3	265	179	68	30	0.2	44	194	2
$f^{\text{extended}}_{\text{recoil}} < 0.1$	1.16±0.05	467	404	20	3.6	188	98	44	17	-	29	26	1
$\frac{3}{4}m_H < m_T < m_H$	1.11±0.10	143	129	14	2.0	59	23	11	11	-	11	14	-

m_T cut, a SOB ratio $\sim 10\%$ can be reached by the whole set of cuts. Besides, the transverse mass m_T is shown in Fig. 5.5(c)(d) [49] for the $e\mu/\mu e$ and $ee/\mu\mu$ channel. In the fit of $e\mu/\mu e$ channel, the same division as $n_j = 0$ channel is applied.

5.3.4 ggF $n_j \geq 2$ channel

To gain as much sensitivity as possible, the ggF signal falling into this channel, which is orthogonal to the VBF $n_j \geq 2$ channel, is also studied, although the expected signal yield is small. Only $e\mu/\mu e$ channels is exploited since the expected sensitivity in $ee/\mu\mu$ channels is too small.

The following cuts are applied after preselection:

- $n_j \geq 2$: to channel-defining cut,
- $n_{bj} = 0$: to reduce top background,
- $m_{\tau\tau} < m_Z - 25\,\text{GeV}$: to reject Drell-Yan $\to \tau\tau$ events,
- Non-VBF selection, to make it orthogonal to the VBF channel by just reversing one of the VBF cuts (which will be explained later in VBF section):
  - VBF cut-based veto: fail of either $\Delta Y_{jj} > 3.6$, $m_{jj} > 600\,\text{GeV}$, CJV or OLV,
  - VBF BDT veto: fail of either CJV, OLV or BDT score > -0.48,

- VH veto, to avoid overlap with VH $n_j \geq 2$ analysis [52]: fail of either $\Delta Y_{jj} < 1.2$ or $|m_{jj} - 85|\,\text{GeV} < 15\,\text{GeV}$, where the value 85 is an average over W and Z boson masses,
- $m_{\ell\ell} < 55\,\text{GeV}$: topological cut which is shown in Fig. 5.6 [49],
- $\Delta\phi_{\ell\ell} < 1.8$: topological cut.

The cutflow corresponds to the above cuts is shown in Table 5.8 [49]. After all the above selection, there are 42 expected signal events over 955 expected background events left. In the signal events, the ggF purity is $\sim 74\%$. From the additional m_T cut, a SOB ratio $\sim 10\%$ can be reached by the whole set of cuts. Besides, the transverse mass m_T is shown in Fig. 5.5(e) [49].

Table 5.8: Cutflow for event selection of the ggF $n_j \geq 2$ category in the 8 TeV data analysis. (see Table 5.6 for presentation details).

Selection	Summary						Composition of N_{bkg}				
	N_{obs}/N_{bkg}	N_{obs}	N_{bkg}	N_{signal}			N_{WW}	N_{top}	N_{misid}	N_{VV}	N_{DY}
				N_{ggF}	N_{VBF}	N_{VH}					
$e\mu/\mu e$ category	0.99±0.00	56759	57180	76	29	24	1330	52020	959	324	2550
$n_{bj} = 0$	1.02±0.01	6777	6650	56	23	15	964	3190	407	233	1850
$m_{\tau\tau} < m_Z - 25$	1.06±0.02	3826	3620	49	19	12	610	2120	248	152	485
VBF orthogonality	1.05±0.02	3736	3550	44	9.0	12	593	2090	241	148	477
VH orthogonality	1.04±0.02	3305	3170	40	8.6	7.4	532	1870	212	132	423
$m_{\ell\ell} < 55$	1.09±0.03	1310	1200	35	7.5	5.0	158	572	124	66	282
$\Delta\phi_{\ell\ell} < 1.8$	1.06±0.03	1017	955	32	6.9	4.5	140	523	99	60	133
$\frac{3}{4}m_H < m_T < m_H$	1.05±0.07	210	200	13.3	2.6	1.9	35	131	16	15	3

5.3.5 VBF channel

As can be seen from the Feynman digram (blue in Fig. 3.7) of the VBF Higgs production, the process is characterised by the presence of two energetic jets with large rapidity gap and no color activity between these jets and the central system, i.e. the Higgs and its decay products [37]. This feature facilitates the design of the event selection. In terms of background, the sources are similar as previous sub-analyses with the top contamination getting stronger due to increased jet multiplicity. Besides, the ggF signal is considered as background here.

There are two sets of event selections: one is a multivariate analysis (MVA) based on boosted decision trees (BDT) [53] and the other is a cut-based analysis, which is used for an independent cross-check of the BDT analysis.

Cut-based analysis

The cut-based analsis is used as a cross-check to the BDT-analysis, especially to check the understanding of the variables used in the BDT training.

The following cuts are applied after preselection:

- $n_j \geq 2$: the channel-defining cut,
- $n_{b\text{-jet}} = 0$: to reduce the top quark background,
- $p_T^{\text{tot}} < 15\,\text{GeV}$: to further suppress $t\bar{t}$ since this process often has larger p_T^{tot} due to the accompanied soft radiation jet which will not be summed up in its calculation,
- $m_{\tau\tau} - m_Z < 25\,\text{GeV}$: to reject Drell-Yan $\to \tau\tau$ events. This cut is applied not only to $e\mu/\mu e$ but also to $ee/\mu\mu$ channels since the Drell-Yan $\to \tau\tau$ is also important there,
- $m_{jj} > 600\,\text{GeV}$: the two leading jets in the signal event tend to be energetic and back-to-back in the forward regions thus leading to a large dijet invariant mass. The m_{jj} distribution is shown in Fig. 5.7(a) [49],
- $\Delta Y_{jj} > 3.6\,\text{GeV}$: for the same reason as above. The ΔY_{jj} distribution is shown in Fig. 5.7(b) [49]. ΔY_{jj} together with the m_{jj} cut are called VBF topological cuts,
- CJV: in the central region defined by the rapidity gap of the two leading jets, low level of hadronic activity is expected in the signal since the mediating weak bosons do not exchange color. This cut helps to suppress backgrounds where jets are produced via QCD radiation,
- OLV: the Higgs boson decay products tend to be lie in the central region. One of the leptons centrality is shown in Fig. 5.7(c) [49],
- $m_{\ell\ell} < 50\,\text{GeV}$: topological cut,
- $\Delta\phi_{\ell\ell} < 1.8$ (2.8) for $p_T^{\text{sublead}} > (<)\ 15\,\text{GeV}$: topological cut.

The cutflow corresponds to the above cuts is shown in Table 5.9 [49]. After all the above selection, the numbers of expected signal events and of background events are comparable, which means that a large SOB ratio can be reached. However, the available statistics is very limited. Besides, the fitting of the m_T distribution after all cuts is done in two m_{jj} bins where the boundary lies at 1 TeV.

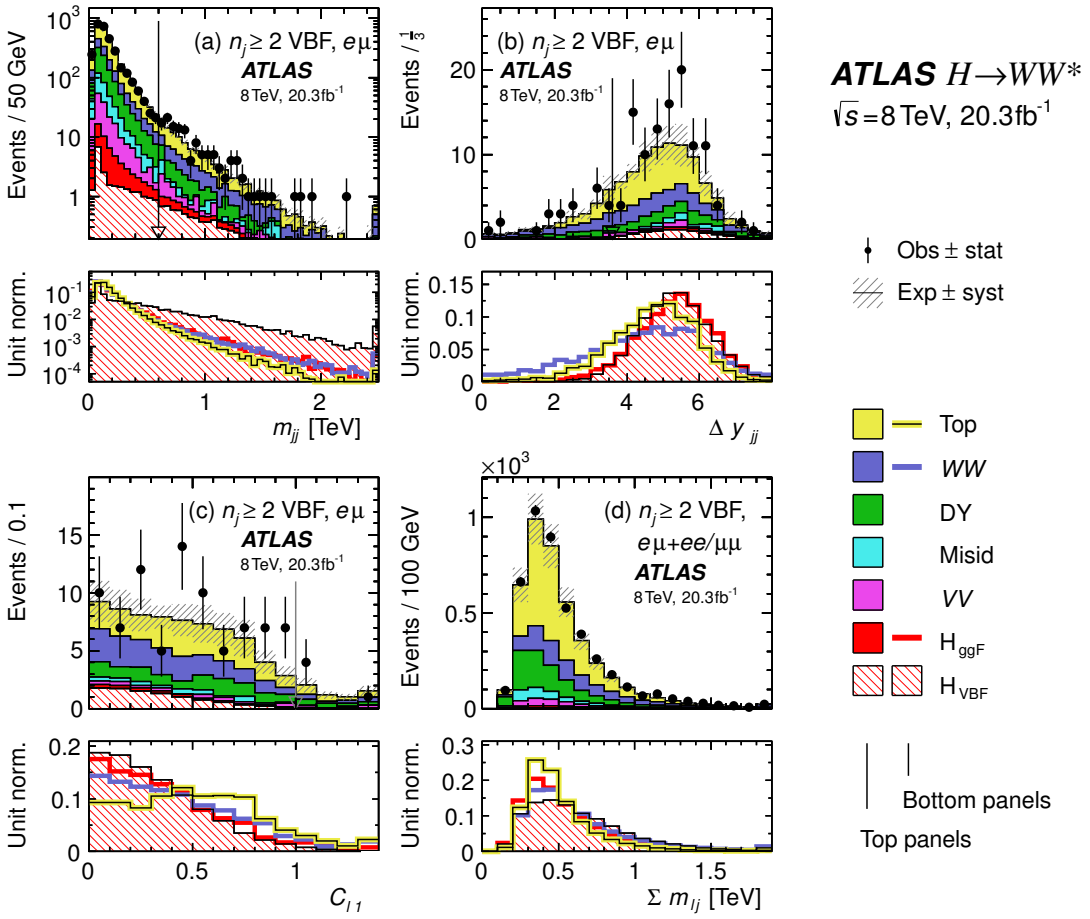


Figure 5.7: Distributions of (a) m_{jj}, (b) ΔY_{jj}, (c) $C_{\ell 1}$, and (d) $\sum_{\ell,j} M_{\ell j}$, for the VBF $n_j \geq 2$ channel. The plot in (a) is made after requiring all selections up to m_{jj}, (b) up to ΔY_{jj} and (c) up to $C_{\ell 1}$. The signal is shown separately for the ggF and VBF production processes. There is no selection made on the variable in (d) since it is only used as an input to the training of the BDT.

Table 5.9: Cutflow for event selection of the VBF $n_j \geq 2$ category in the 8 TeV data analysis. (see Table 5.6 for presentation details). The expected yields for WW and Drell-Yan $\rightarrow \tau\tau$ are divided into QCD and electroweak (EW) processes.

Selection	Summary						Composition of N_{bkg}									
	N_{obs}/N_{bkg}	N_{obs}	N_{bkg}	N_{signal}			N_{WW}		N_{top}		N_{misid}		N_{VV}	N_{DY}		
				N_{ggF}	N_{VBF}	N_{VH}	N_{WW}^{qcd}	N_{WW}^{ew}	$N_{t\bar{t}}$	N_{st}	N_{W+jets}	N_{QCD}		$N_{ee/\mu\mu}$	$N_{\tau\tau}^{qcd}$	$N_{\tau\tau}^{ew}$
$e\mu/\mu e$ sample	1.00±0.00	61434	61180	85	32	26	1350	68	51810	2970	847	308	380	51	3260	46
$n_{bj} = 0$	1.02±0.01	7818	7700	63	26	16	993	43	3000	367	313	193	273	35	2400	29
$p_T^{tot} < 15$	1.03±0.01	5787	5630	46	23	13	781	38	1910	270	216	107	201	27	2010	23
$m_{\tau\tau} - m_Z < 25$	1.05±0.02	3129	2970	40	20	9.9	484	22	1270	177	141	66	132	7.6	627	5.8
$m_{jj} > 600$	1.31±0.12	131	100	2.3	8.2	-	18	8.9	40	5.3	1.8	2.4	5.1	0.1	15	1.0
$\Delta Y_{jj} > 3.6$	1.33±0.13	107	80	2.1	7.9	-	11.7	6.9	35	5.0	1.6	2.3	3.3	-	11.6	0.8
CJV	1.36±0.18	58	43	1.3	6.6	-	6.9	5.6	14	3.0	1.3	1.3	2.0	-	6.8	0.6
OLV	1.42±0.20	51	36	1.2	6.4	-	5.9	5.2	10.8	2.5	1.3	1.3	1.6	-	5.7	0.6
$m_{\ell\ell}, \Delta\phi_{\ell\ell}, m_T$	2.53±0.71	14	5.5	0.8	4.7	-	1.0	0.5	1.1	0.3	0.3	0.3	0.6	-	0.5	0.2
$ee/\mu\mu$ sample	0.99±0.01	26949	27190	31	14	10.1	594	37	23440	1320	230	8.6	137	690	679	16
$n_{bj}, p_T^{tot}, m_{\tau\tau}$	1.03±0.03	1344	1310	13	8.0	4.0	229	12.0	633	86	26	0.9	45	187	76	1.5
$m_{jj}, \Delta Y_{jj}, CJV, OLV$	1.39±0.28	26	19	0.4	2.9	0.0	3.1	3.1	5.5	1.0	0.2	0.0	0.7	3.8	0.7	0.1
$\ell\ell, \Delta\phi_{\ell\ell}, m_T$	1.63±0.69	6	3.7	0.3	2.2	0.0	0.4	0.2	0.6	0.2	0.2	0.0	0.1	1.5	0.3	0.1

BDT analysis

The BDT analysis is the baseline method for the VBF channel with the BDT defined as following: A decision tree is a collection of cuts designed to classify events as signal-like or background-like. A given signal event is correctly identified if it is placed in a signal-dominated leaf, and vice-versa for background events. After the initial tree is built, another tree is grown to better separate the signal and background events that were misidentified by the first tree. This proceeds iteratively until there is a collection of a specified number of trees, in a process known as boosting. A weighted average is taken from all these trees to form a BDT output discriminant with values ranging between -1 and 1. The events with weights close to $+1$ (-1) are signal (background) like.

The BDT training variables are chosen by employing an "N$-$1 minimal loss variable pruning" procedure [37]:

- Start with a BDT trained with a maximal set of potentially useful discriminating variables,
- Remove one variable at a time and evaluate the performance of the $(N-1)$-variable BDT based on expected significance (using only statistical uncertainties),
- Take the best performing $(N-1)$-variable BDT as the new benchmark.

Finally eight variables are selected: ΔY_{jj}, m_{jj}, and η_{lep} centrality, which exploit the VBF topology; $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, and m_T, which are sensitive to the spin correlation topology; p_T^{tot} and $\sum_{\ell,j} M_{\ell j}$, which are mainly for the rejection of $t\bar{t}$.

With the BDT variables defined, some of the event selections used in the cut-based analysis, which are the cuts on n_{bj}, $m_{\tau\tau}$, CJV and OLV, are applied also to the BDT analysis.

Unlike the ggF dedicated analysis, the output BDT instead of m_T will be used in the final fit as shown in Fig. 5.8(a)(b) [49]. It will be binned into four categories, of which the three high BDT score bins will be fitted (among them, the last bin has the maximum sensitivity).

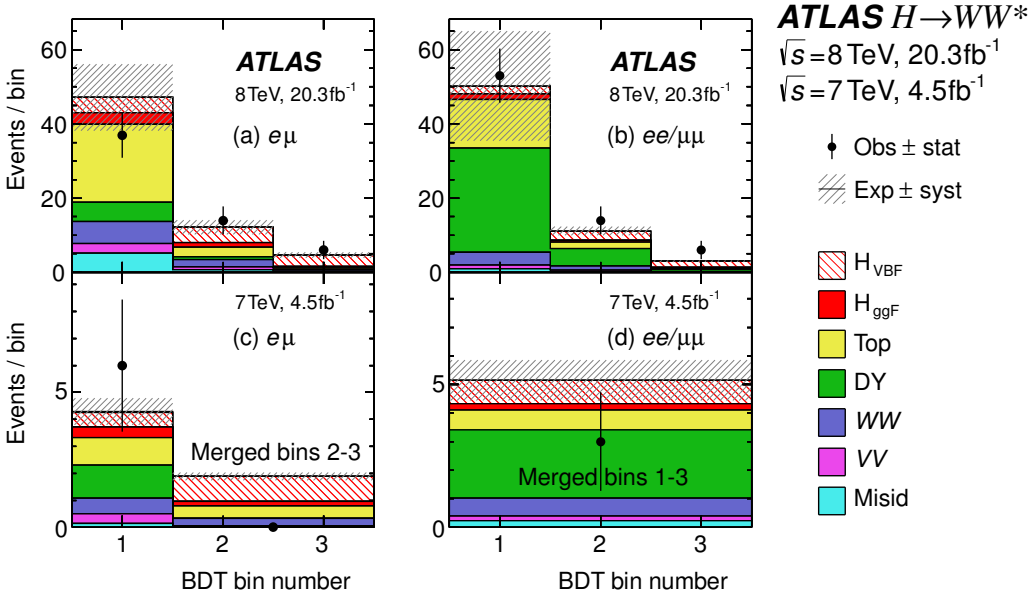


Figure 5.8: Distributions of the BDT output in the VBF $n_j \geq 2$ category in the 8 and 7 TeV data analyses. The plot is made after requiring all the selections listed in Table 5.10 and after the BDT binning.

5.3.6 7 TeV analysis

The 7 TeV data collected in year 2011 is re-analysed in a similar way as the above. The motivation is to incorporate the improvements made in the 8 TeV analysis and make the event selection as well as samples compatible with those of 8 TeV analysis to make the combination of these two analyses easier. It is also divided into same flavour and different flavour channels as well as different jet multiplicity channels, with the exception that the $n_j \geq 2$ channel is only used for the VBF analysis.

The object definitions are slightly different from those of the 8 TeV analysis. Different triggers are used as already shown in Sec. 4.2. Different reconstruction

and identification strategies are applied for the electron, e.g. no likelihood electron identification method in the 7 TeV analysis. Different collections of jets as well as different pile-up suppression cuts are used: EM scale jets are selected and the JVF cut altered in the 7 TeV analysis.

The E_T^{miss} cut of the common preselection is changed (lowered) due to the lower pile-up condition in the 7 TeV data:

- For $n_j = 0$ or 1 channel: $E_T^{\text{miss, track, jetCorr}} > 20\,\text{GeV}$ is used in $e\mu/\mu e$ while $E_{T,\text{Rel}}^{\text{miss, calo}} > 35\,\text{GeV}$ is used in $ee/\mu\mu$,
- For VBF channel: $E_T^{\text{miss, calo}} > 45\,\text{GeV}$ is used in $ee/\mu\mu$ while no cut in $e\mu/\mu e$.

The event selections in the $e\mu/\mu e$ channels of $n_j = 0$ or 1 are exactly the same as in the 8 TeV analysis. But those in the $ee/\mu\mu$ channels are slightly modified:

- In the $n_j = 0$ and $ee/\mu\mu$ channels:
  - $p_{T,\ell\ell}$ cut is raised to 40 GeV,
  - $E_{T,\text{Rel}}^{\text{miss, track}}$ cut is removed,
  - f_{recoil} cut loosened to 0.2.
- In the $n_j = 1$ and $ee/\mu\mu$ channels:
  - $E_{T,\text{Rel}}^{\text{miss, track}}$ cut is replaced by $p_T^{\text{tot}} > 35\,\text{GeV}$,
  - f_{recoil} cut loosened to 0.5.

As in the 8 TeV case, the same m_T is used in the final fit, whose distributions in each $n_j \leq 1$ channels are shown in Fig. 5.9 [49].

For the 7 TeV VBF analysis, the BDT trained using 8 TeV samples are applied to benefit from the high statistics in the 8 TeV dataset. Besides, by using the same BDT definition, the same theoretical uncertainty is shared, facilitating the combination. As for the event selections, no change is made for the BDT-based analysis and the BDT output used for the fit is shown in Fig. 5.8(a)(b) [49]. While the same but slightly loosened cuts are used for the cut-based analysis.

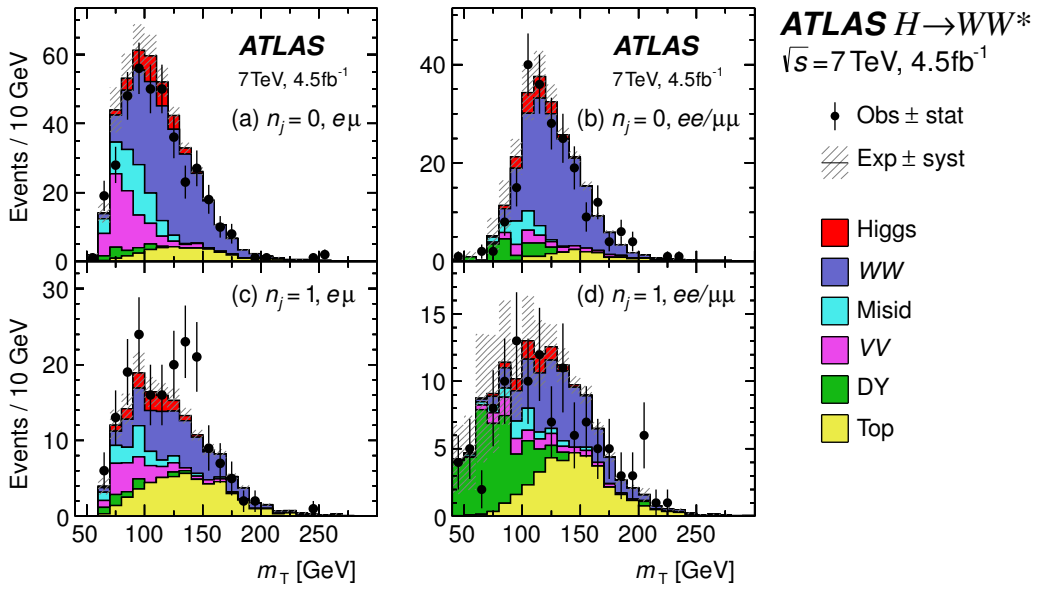


Figure 5.9: Distributions of the transverse mass m_T for the $n_j \leq 1$ categories in the 7 TeV data analysis. The plot is made after requiring all selections up to m_T.

5.3.7 Summary

All the above event selections (including triggers) identify various signal regions (SRs) out of the basic event collections where all kinds of processes are not filtered. By applying these selections, the signal sensitivity, to some extent can be revealed by the SOB ratio, grows vastly. As mentioned in the section introducing the pp collision, the Higgs signal is estimated to be ten or eleven orders of magnitude smaller than the total pp cross section, which corresponds to an extremely tiny SOB value. However, if looking at the end of the above cutflows, the SOB value is expected to be in the order of 10^{-1} after all these selections, from which the feat of the analysis can be seen.

On one side, the various backgrounds are greatly suppressed. On the other side, the signals are retained with a significant efficiency. In the 8 TeV analysis, including all signal categories and production modes, 10.2% of the $H \to WW^* \to \ell\nu\ell\nu$ events are selected. When including only VBF production mode, the number is 7.8% [49].

At last, all the event selections related to 8 TeV analysis are summarised in Table 5.10 [49].

Table 5.10: Event selection summary for the analysis of the 8 TeV data for $m_H = 125$ GeV; Selection requirements specific to the $e\mu/\mu e$ and $ee/\mu\mu$ lepton-flavour samples are noted as such (otherwise, they apply to both); a dash (-) indicates no selection. All energy-related values are in GeV.

Objective		ggF			VBF
		$n_j = 0$	$n_j = 1$	ggF $n_j \geq 2$	VBF $n_j \geq 2$
Preselection	All n_j	$p_T^{lead} > 22$ for the leading lepton $p_T^{sublead} > 10$ for the sub-leading lepton Opposite-charge leptons $m_{\ell\ell} > 10$ for the $e\mu/\mu e$ sample $m_{\ell\ell} > 12$ for the $ee/\mu\mu$ sample $\lvert m_{\ell\ell} - m_Z \rvert > 15$ for the $ee/\mu\mu$ sample			
		$E_T^{\text{miss,track,jetCorr}} > 20$ for $e\mu/\mu e$	$E_T^{\text{miss,track,jetCorr}} > 20$ for $e\mu/\mu e$	$E_T^{\text{miss,track,jetCorr}} > 20$ for $e\mu/\mu e$	No E_T^{miss} requirement for $e\mu/\mu e$
		$E_{T,Rel}^{\text{miss,calo}} > 40$ for $ee/\mu\mu$	$E_{T,Rel}^{\text{miss,calo}} > 40$ for $ee/\mu\mu$	-	-
Reject backgrounds	DY	$E_{T,Rel}^{\text{miss,track}} > 40$ for $ee/\mu\mu$	$E_{T,Rel}^{\text{miss,track}} > 35$ for $ee/\mu\mu$	-	$E_T^{\text{miss,track,jetCorr}} > 40$ for $ee/\mu\mu$
		$f_{recoil} < 0.1$ for $ee/\mu\mu$	$f_{recoil} < 0.1$ for $ee/\mu\mu$	-	$E_T^{\text{miss,calo}} > 45$ for $ee/\mu\mu$
		$p_{T,\ell\ell} > 30$	$m_{\tau\tau} < m_Z - 25$	$m_{\tau\tau} < m_Z - 25$	$m_{\tau\tau} < m_Z - 25$
		$\Delta\phi_{\ell\ell,E_T^{\text{miss}}} > \pi/2$	-	-	-
	Misid.	-	$M_{T,\max}^W > 50$ for $e\mu/\mu e$	-	-
	Top	$n_j = 0$	$n_{bj} = 0$	$n_{bj} = 0$	$n_{bj} = 0$
		-	-	-	p_T^{tot} inputs to BDT
		-	-	-	$\sum_{\ell,j} M_{\ell j}$ inputs to BDT
VBF topology		-	-	See Sec. 5.3.4 for rejection of VBF & VH ($W, Z \to jj$),	m_{jj} inputs to BDT ΔY_{jj} inputs to BDT $\sum C_l$ inputs to BDT $C_{\ell 1} < 1$ and $C_{\ell 2} < 1$ $C_{j_3} > 1$ for j_3 with $p_T^{j_3} > 20$ BDT ≥ -0.48
Higgs decay topology		$m_{\ell\ell} < 55$ $\Delta\phi_{\ell\ell} < 1.8$ No m_T requirement	$m_{\ell\ell} < 55$ $\Delta\phi_{\ell\ell} < 1.8$ No m_T requirement	$m_{\ell\ell} < 55$ $\Delta\phi_{\ell\ell} < 1.8$ No m_T requirement	$m_{\ell\ell}$ inputs to BDT $\Delta\phi_{\ell\ell}$ inputs to BDT m_T inputs to BDT

5.4 Background estimation

The basic methodology of background estimation in the analysis is to normalise the background or get direct data-driven estimation making use of control regions (CR) which are usually defined by inverting some of the event selection criteria to make it orthogonal to the signal region and by loosening or dropping some cuts to increase statistics of the CR. But in some cases, pure MC prediction or pure MC plus validation region (VR) will be used. Besides, all the methods are described in terms of the 8 TeV analysis since almost the same procedures are exported to the 7 TeV analysis. It will be described when different methods are used.

The data-driven estimation or control region method can be simply expressed in terms the following formula:

$$N_{SR}^{\text{estimated}} = N_{\text{SR}}^{\text{MC}} \cdot N_{\text{CR}}^{\text{data}} / N_{\text{CR}}^{\text{MC}} = N_{\text{SR}}^{\text{MC}} \cdot \beta \quad (5.7)$$

$$= N_{\text{CR}}^{\text{data}} \cdot N_{\text{SR}}^{\text{MC}} / N_{\text{CR}}^{\text{MC}} = N_{\text{CR}}^{\text{data}} \cdot \alpha, \quad (5.8)$$

where $N_{\text{SR}}^{\text{MC}}$ and $N_{\text{CR}}^{\text{MC}}$ are the MC prediction of the background process in the signal and control regions, $N_{\text{CR}}^{\text{data}}$ is the observed data in the control region, and the $N_{\text{SR}}^{\text{estimated}}$ is the estimated background yield in the signal region. Besides, the

Table 5.11: Summary of background treatments for the 8 TeV analyses. The estimation procedures for various background processes are given in four categories: normalised using a control region (CR); data-derived estimate (Data); normalised using the MC (MC); and normalised using the MC, but validated in a control region (MC + VR). The "($e\mu/\mu e$)" terms denote that for the $ee/\mu\mu$ channel in the same jet bins , the $e\mu/\mu e$ region is used instead, for reasons of purity and/or statistics. The "(all)" terms denote that the control region combines all four lepton flavour channels.

Channel	WW	Top	$Z/\gamma^* \to \tau\tau$	$Z/\gamma^* \to ee/\mu\mu$	Wjets/QCD	VV
$n_j = 0$						
$e\mu/\mu e$	CR	CR	CR	MC	Data	CR
$ee/\mu\mu$	CR ($e\mu/\mu e$)	CR ($e\mu/\mu e$)	CR ($e\mu/\mu e$)	Data	Data	MC + VR
$n_j = 1$						
$e\mu/\mu e$	CR	CR	CR	MC	Data	CR
$ee/\mu\mu$	CR ($e\mu/\mu e$)	CR ($e\mu/\mu e$)	CR ($e\mu/\mu e$)	Data	Data	MC + VR
ggF $n_j \geq 2$						
$e\mu/\mu e$	MC	CR	CR	MC	Data	MC + VR
VBF $n_j \geq 2$						
$e\mu/\mu e$	MC+VR	CR (all)	CR (all)	MC	Data	MC + VR
$ee/\mu\mu$	MC+VR	CR (all)	CR (all)	Data	Data	MC + VR

extrapolation factor α and the normalisation factor β are defined, each revealing the essence of a type of the background estimation method: the method making use of α will have the signal region and control region simultaneously into the final fit, e.g. the WW, while the other type relevant to β will have only the normalised yield in signal region entering the fit, e.g. the W+jets and QCD.

Before introducing the background estimation process by process, an overview of the methods to be used is given in Table 5.11 [37].

5.4.1 WW

WW is the most important background in $n_j = 0$ and $n_j = 1$ channels and enters all of the analysis channels since it has exactly the same decay products as the signal process, except that it is not from a resonance. Thus data-driven methods are indispensable. As shown in Table 5.11, the WW CRs are defined in $n_j \leq 1$ channels. The number of WW events in the $n_j = 0$ or 1 channel $N^{\text{estimated}}_{WW,\text{SR}\,i}$

is estimated using Formula 5.8 as:

$$N^{\text{estimated}}_{WW,\text{SR}\,i} = N^{MC}_{WW,\text{SR}\,i} \cdot \frac{N^{\text{data}}_{WW,\text{CR}\,i}}{N^{\text{MC}}_{WW,\text{CR}\,i}} = N^{MC}_{WW,\text{SR}\,i} \cdot \beta_i, \tag{5.9}$$

$$= N^{\text{data}}_{WW,\text{CR}\,i} \cdot \frac{N^{MC}_{WW,\text{CR}\,i}}{N^{\text{MC}}_{WW,\text{CR}\,i}} = N^{\text{data}}_{WW,\text{CR}\,i} \cdot \alpha_i\,, \tag{5.10}$$

where "i" takes the value of 0 or 1, representing the jet bin and the meaning of each number is obvious by reading their supper-/sub-scripts. The α version of the formula is used in the final fit since it has the advantage that most of systematic uncertainties enter into this factor. While using the β version, the normalisation factors can be calculated to have a sense of the level of agreement between data and MC modelling.

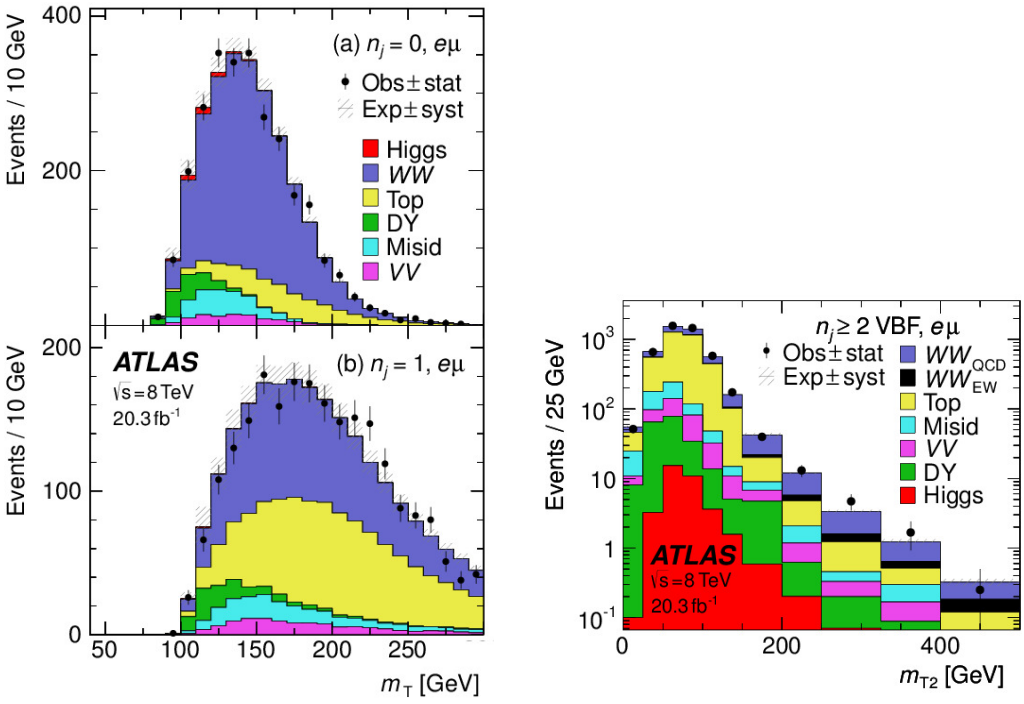
As for the $n_j \geq 2$ channels, due to the large top quark background contamination (for VBF) or limited importance (for ggF), the WW process is estimated from Sherpa MC prediction and normalised to the NLO calculation from MCFM [54]. Besides, the WW pair production from the DPI is also investigated and found to be very small thus the MC based prediction is used.

WW in the $n_j = 0$ channel

The WW control region in this channel is defined after the $p_{T,\ell\ell} > 30\,\text{GeV}$ cut of the $n_j = 0$ signal region selection by the following requirements:

- $p_T^{\text{sublead}} > 15\,\text{GeV}$: to raise sub-leading lepton p_T cut from 10 to 15 GeV for the suppressing of W+jets contamination. Besides, WW in the 10–15 GeV bin is negligible and has low purity,
- $\Delta\phi_{\ell\ell} < 2.6$: to reduce $Z/\gamma^* \to \tau\tau$ contamination,
- $55 < m_{\ell\ell} < 110\,\text{GeV}$: where the lower bound is exactly the invert of the signal region cut on $m_{\ell\ell}$ to make the WW control region orthogonal to the SR. Actually this 55 GeV boundary is a compromise of the signal acceptance, statistical power of the CR, while ensuring similar kinematics in the SR and CR. The upper bound is added for reducing the theoretical uncertainty of the extrapolation from the control region to the SR.

The control region selected by the above cuts is dominated by the WW process and has a purity of $\sim 70\%$. The m_T distribution in this region is shown in the



(a) $n_j \leq 1$ (b) VBF $N_{ket} \geq 2$

Figure 5.10: WW control region distributions of transverse mass m_T for the $n_j \leq 1$ channels and the WW validation region distributions of the m_{T2} variable for the VBF $n_j \geq 2$ channel.

upper plot of Fig. 5.10(a) [49]. The normalisation factor measured from this control region is 1.22 ± 0.03 (stat.).

The $n_j = 0$ WW control region is only defined in the $e\mu/\mu e$ channel since there is a significant Drell-Yan contamination in the WW control region for the $ee/\mu\mu$ channel. So the normalisation factor derived in this CR is applied to all lepton flavour channels.

Besides, there is a WW validation region which does not enter the fit but just to make consistency check. It is defined just by replacing the original $m_{\ell\ell}$ cut of the CR with $m_{\ell\ell} > 110\,\text{GeV}$. The normalisation factors derived from the CR and VR are compared and found to be consistent at the level of $1.1\,\sigma$.

WW in the $n_j = 1$ channel

The WW control region in the $n_j = 1$ case is similarly defined after the $M^W_{T,\text{max}} > 50\,\text{GeV}$ cut of the $n_j = 1$ signal region selection by the following requirements:

- $p_T^{\text{sublead}} > 15\,\text{GeV}$: motivated by the same reason as in the case of $n_j = 0$ control region,
- $|m_{\tau\tau} - m_Z| > 25\,\text{GeV}$: to reduce $Z/\gamma^* \to \tau\tau$ contamination and should be distinguished from the similar cut used in the signal region selection $m_{\tau\tau} < m_Z - 25$. The $m_{\tau\tau}$ cut here is different from that used in the signal region definition, which is $m_{\tau\tau} < m_Z - 25\,\text{GeV}$. This is because if using the latter, we will lose 30% of the WW events in the high $m_{\tau\tau}$ tail,
- $m_{\ell\ell} > 80\,\text{GeV}$: to move away from the signal region as well as further reject $Z/\gamma^* \to \tau\tau$ contamination.

The control region selected by the above cuts is dominated by both WW and top processes and has a WW purity of $\sim 40\%$. Thus the top event contribution should be estimated first before the WW estimation. The m_T distribution in this region is shown in the bottom plot of Fig. 5.10(a) [49]. The normalisation factor measured from this control region is 1.05 ± 0.05 (stat.). Similar to the $n_j = 0$ case, the WW CR in the $n_j = 1$ channel is also defined in the $e\mu/\mu e$ channel but applied to all lepton flavour channels.

WW in the $n_j \geq 2$ channel

Due to the limited WW background contribution in the ggF dedicated channel and the large contamination from the top quark background to the construction of the WW control region in the VBF dedicated channel, the MC prediction is used for simplicity. The WW process considered here is generated by Sherpa and includes Feynman diagrams containing both QCD and electroweak vertices, which are called "QCD WW+2jets" and "EW WW+2jets", respectively.

In the VBF dedicated channel, the "QCD WW+2jets" sample is checked in a validation region motivated by the possible large theoretical uncertainty associated with this sample, of which the details can be found in [55]. The key selection to define this validation region is a m_{T2} cut [56], whose distribution is shown in Fig. 5.10(b) [49]. The purity of the VR reaches $\sim 60\%$ and the ratio of data over MC in this region is 1.15 ± 0.19(stat.), which means that the data and MC agree with each other within the statistical error.

WW in 7 TeV

WW background estimation in the 7 TeV analysis following exactly the same procedures described above.

5.4.2 Top

The top quark background corresponds to those processes that include at least one top quark production. At the LHC, top quarks are produced either in pairs ($t\bar{t}$) or in association with a W boson (Wt) or quarks (other single tops). The top decay gives one W boson and one *bottom* quark which finally developed as b-jet. So the $t\bar{t}$ or Wt becomes background when the b-jet is not identified by the b-tagger, with a signature of two W bosons (which subsequently decay into leptons) plus jets (one or two b-jets and possibly more radiation jets). The other single top processes comes into play by a different mechanism: except for the missed b-jet by the tagger, the quark associated with the top should be reconstructed as a lepton, or else only one lepton in the final state which will be filtered at the beginning of the common preselection. However, these processes are of very tiny amount and estimated together with $t\bar{t}$ and Wt for simplicity. Different data-driven methods are applied for the estimation of the top backgrounds in each jet multiplicity channels, which will be introduced in detail in the following.

Top in the $n_j = 0$ channel

Top backgrounds cause contamination in this channel when all of its jets (b-jet or radiation jet) are too soft to pass the good jet selection criteria. This contamination at the cut stage before fit is small compared to the other dominating backgrounds in the $n_j = 0$ channel, however still comparable to the expected signal contribution. Thus a data-driven method called Jet Veto Survival Probability (JVSP) [57] is developed for its estimation which is performed in the $e\mu/\mu e$ channel and applied to all lepton flavour channels. I have been responsible for the top background estimation as well as the systematics assessments in this particular channel. As a reminder, the b-jet used in this section is defined with a p_T threshold of 25 GeV to be the same as the jet multiplicity definition. In the other channels, a lower threshold of 20 GeV is used to suppress the larger top quark contribution.

The jet veto survival probability of the top processes is defined as:

$$P_2 = \frac{N_{\text{top},0j}}{N_{\text{top, inc}}}\,, \tag{5.11}$$

where $N_{\text{top},0j}$ is the number of the top background falling into the $n_j = 0$ channel and $N_{\text{top, inc}}$ is the number of inclusive top background on which no jet selection is imposed. P_2 is just a convenient choice of denotation of the JVSP, which will be clear later. The meaning of P_2 is illustrated in Fig. 5.1, which is the fraction of top events (yellow histograms) in the $n_j = 0$ bin.

So $N_{\text{top},0j}$ could be calculated by:

$$N_{\text{top},0j} = N_{\text{top, inc}} \cdot P_2\,. \tag{5.12}$$

The above formula is not just a matter of re-arrangement of the original Formula 5.11. It inspires us to decompose the task of the $N_{\text{top},0j}$ estimation into the $N_{\text{top, inc}}$ and P_2 estimation. This equation is similar as the general Formula 5.8, with the α replaced by P_2.

The data-driven estimation of $N_{\text{top, inc}}$ is easy since after the common preselection the top quark background dominants the data sample.Thus the estimated inclusive top sample is calculated by subtracting from data the non-top events using either MC or data-driven estimation:

$$N^{\text{estimated}}_{\text{top, inc}} = N^{\text{data}}_{\text{inc}} - N_{\text{non-top, inc}}. \tag{5.13}$$

In practice, the inclusive region is defined by modifying the E_T^{miss} cut in the common preselection from a jet-bin dependent cut to a universal cut of $E_T^{\text{miss, track, jetCorr}} > 20\,\text{GeV}$ and adding an additional $\Delta\phi_{\ell\ell} < 2.8$ cut to reduce the $Z/\gamma^* \to \tau\tau$ contamination. The resulting region is dominated by top events with a purity of 74%.

The P_2 is corrected using a data-driven method to account for possible bias from mis-modelling of jets (b-jet or radiation jet). Denote P_1 as the probability of a b-jet not being tagged, either due to not passing the jet selection or inefficiency of the tagger algorithm, and P_0 as the probability of the other jets passing the jet veto cut (not the jet selection cut), including both radiation jet and untagged b-jet. Then the JVSP can be expressed as:

$$P_2 = P_0 \cdot P_1^2. \tag{5.14}$$

The P_1^2 emerges from the assumption that two b-jets in the top event are independently being tagged or not. This independence is checked using two subsamples of the inclusive top quark MC sample, one with only one b-jet being tagged. And the other with both b-jets being tagged. So we have the equations:

$$N_{s1} = N_{\text{top, inc}} \cdot P_1(1 - P_1), \tag{5.15}$$

$$N_{s2} = N_{\text{top, inc}} \cdot P_1^2, \tag{5.16}$$

where N_{s1} and N_{s2} are the numbers of events of the two subsamples, and $N_{\text{top, inc}}$ the number of events of the inclusive top sample. The P_1 values calculated from the above two subsamples are found to agree within 2%, which supports the independence consideration.

Now the task of correcting P_2 breaks down to that of corrections P_0 and P_1^2. The relative importance of the mis-modelling in P_0 and P_1^2 can be assessed by comparing the multiplicity distribution of tagged b-jets and the rest jets, which is shown in Fig. 5.11. The black histogram represents for the tagged b-jets and the red dashed histogram for the other jets. From the plot, the $n_j = 0$ bin fraction of the tagged b-jets (P_1^2) is much smaller than that of the others (P_0). So the jet bin migration of P_0 is expected to be much smaller than that of P_1^2 due to various possible systematic effects. Thus the mis-modelling of P_0 can be neglected and the correction factor for P_2 can be defined as:

$$\gamma_2 = \frac{P_2^{\text{estimated}}}{P_2^{MC}} \simeq \left(\frac{P_1^{\text{estimated}}}{P_1^{\text{MC}}} \right)^2 = \gamma_1, \tag{5.17}$$

where γ_1 and γ_2 are the corrections to P_1 and P_2, respectively.

To get the correction to P_1, a top control region is defined by requiring $n_{b\text{-jet}} \neq 0$ from the inclusive sample defined in calculating $N_{\text{top, inc}}^{\text{estimated}}$. The number of event of this CR is denoted as $N_{\text{top, CR}}$. In this control sample, a tagged b-jet is chosen (one of them is chosen randomly if there is more than one b-jet), the number of probing jet $N_{\text{top, CR}}^{0\,\text{prob}}$ is calculated by requiring the distance ΔR between the probing jet and the tagged b-jet be greater than 1. In the simplest case where there is no radiation jet and the two b-jets are back-to-back (thus $\Delta R > 1$ cut is satisfied) in the event, the probing jet is just the other b-jet.$N_{\text{top, CR}}$ is actually the sum of N_{s1} and N_{s2},

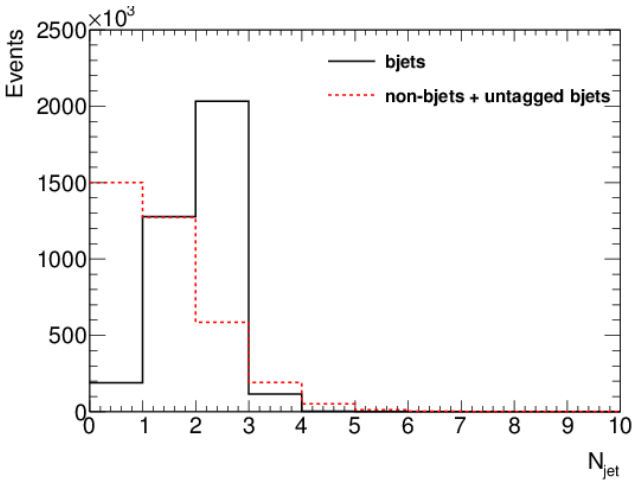


Figure 5.11: The jet multiplicity distribution in the inclusive $t\bar{t}$ MC sample for the b-jets and other non-b-jets and untagged b-jets

and $N^{0\,\mathrm{prob}}_{\mathrm{top,\,CR}} = N_{s2}$. Thus P_1 can be calculated making use of Formula 5.16 as:

$$\begin{aligned} \frac{N^{0\,\mathrm{prob}}_{\mathrm{top,\,CR}}}{N_{\mathrm{top,\,CR}}} &= \frac{N_{s2}}{N_{s1} + N_{s2}} \\ &= \frac{N_{\mathrm{top,\,inc}} \cdot P_1^2}{N_{\mathrm{top,\,inc}} \cdot P_1^2 + N_{\mathrm{top,\,inc}} \cdot P_1(1 - P_1)} \\ &= P_1 \,. \end{aligned} \tag{5.18}$$

The $\Delta R > 1$ cut is defined to suppress the effect of counting radiation jets accompanied the tagged jet in the calculation of P_1.The probing jet being selected from the jet collection rather than the b-jet collection makes the method robust against the performance of the b-jet tagger.

Relavant distributions in the control regions are checked by comparing kinematic distributions of the tagging b-jet and the probing jet, which is shown in Fig. 5.12. Good agreement between data and MC predictions is observed. These top control regions are of high purity.

Put together all pieces, we get the expression for the estimated top background

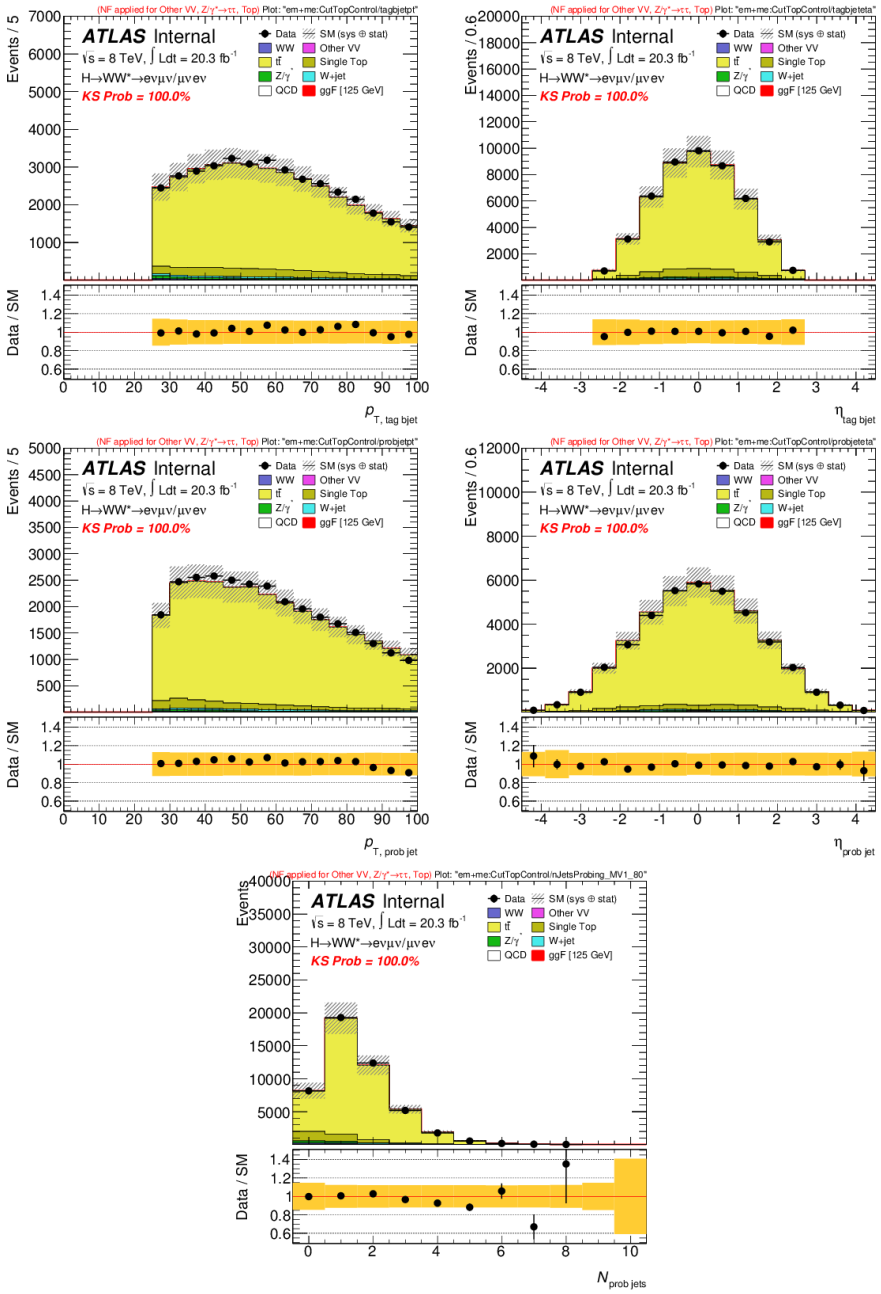


Figure 5.12: Jet distributions in data in $e\mu$ and μe channel compared to MC in the $n_{b\text{-jet}} \neq 0$ control sample used to calculating P_1 for the top quark background estimate. The upper plots display the transverse momentum (top-left) and pseudo-rapidity (top-right) of the tagging b-jets. The other plots display the transverse momentum (middle-left) and pseudo-rapidity (middle-right) and the multiplicity of the probing jets (bottom). The calculation of the KS probabilities considers only the statistical uncertainties, the hashed band is showing the systematic uncertainties on the background prediction.

in the $n_j = 0$ channel out of the prototype in Formula 5.12:

$$\begin{aligned} N_{\text{top},0j}^{\text{estimated}} &= N_{\text{top, inc}}^{\text{estimated}} \cdot P_2^{\text{estimated}} \\ &= N_{\text{top, inc}}^{\text{estimated}} \cdot P_2^{\text{MC}} \cdot \left(\frac{P_1^{\text{estimated}}}{P_1^{\text{MC}}} \right)^2 \\ &= N_{\text{top, inc}}^{\text{estimated}} \cdot \left(P_1^{\text{estimated}} \right)^2 \cdot \frac{P_2^{\text{MC}}}{\left(P_1^{\text{MC}} \right)^2} \,. \end{aligned}$$

From the above expression, it can be seen that all the possible top MC mismodelling lies in the term $\frac{P_2^{\text{MC}}}{(P_1^{\text{MC}})^2}$ which can be used to assess the theoretical and experimental systematic uncertainties of the JVSP method. Besides, due to the high purity and large statistics of the top control regions involved, the non-top background contamination is very limited and the statistical uncertainty of the method is small. By dividing $N_{\text{top},0j}^{\text{estimated}}$ with $N_{\text{top},0j}^{\text{MC}}$, the normalisation factor can be calculated to be: 1.08 ± 0.02 (stat.).

The above JVSP method is carried out at the jet veto cut level of the $n_j = 0$ signal region selection, thus predicts only top events at that cut stage. To estimate the top backgrounds in the final signal region, this estimated number of top events $N_{\text{top},0j}^{\text{estimated}}$ needs to be extrapolated from the jet veto cut to the final signal region $N_{\text{top, SR}}$ (after the $\Delta\phi_{\ell\ell}$ cut) with the extrapolation factor:

$$\alpha_{0j \to \text{SR}} = \frac{N_{\text{top, SR}}}{N_{\text{top},0j}} \,, \tag{5.19}$$

which is estimated by MC, bringing in additional extrapolation uncertainty. Moreover, as the top has a sizeable contribution to the WW $n_j = 0$ control region, the estimated top at the jet veto cut is also extrapolated to the WW CR by the MC predicted extrapolation factor:

$$\alpha_{0j \to WW\,\text{CR}} = \frac{N_{\text{top, } WW\text{ SR}}}{N_{\text{top},0j}} \,. \tag{5.20}$$

There are several other competing methods for the $n_j = 0$ top estimation which will be discussed in detail in Appendix A. The conclusion is that the JVSP method performs best. Besides, the comparison is also carried out in other jet multiplicity bins.

Top in the $n_j = 1$ channel

The top process is the second largest background in this channel which amounts to $\sim 30\%$ of the total backgrounds. It falls into the $n_j = 1$ category when only one of its final state jets is identified as good jet. A large portion of these events can be removed by imposing the b-jet veto cut $n_{b\text{-jet}} = 0$ which is already introduced in the event selection section. There will still be some top events passing this cut due to the inefficiency of the b-tagger. The data-driven method to be introduced here, called in-situ b-tagging efficiency method, is to estimate the top background right after this cut. As before, it is also implemented in the $e\mu/\mu e$ channel and applies its results to all lepton flavour channels.

The basic formula to start from is the following:

$$\epsilon_{\text{top, tag}} = \frac{N_{\text{top, tag}}}{N_{\text{top, all}}}\,, \tag{5.21}$$

which is the definition of the top tagging efficiency $\epsilon_{\text{top, tag}}$. $N_{\text{top, tag}}$ is the number of top events tagged as top. In the real situation, it is the total number of events in the region which is defined by applying all the $n_j = 1$ channel selection up to the b-jet veto and then reversing this cut to be $n_{b\text{-jet}} \neq 0$. $N_{\text{top, all}}$ is the number of top events to be tagged which corresponds to the region, in practice, defined by applying all $n_j = 1$ channel cut up to the b-jet veto with the b-jet veto excluded. So the number of top events after the b-jet veto cut can be expressed as:

$$N_{\text{top, untag}} = \frac{N_{\text{top, tag}}}{\epsilon_{\text{top, tag}}} \cdot (1 - \epsilon_{\text{top, tag}})\,. \tag{5.22}$$

The estimation of $N_{\text{top, untag}}$ decomposes into that of $N_{\text{top, tag}}$ and the top tag efficiency $\epsilon_{\text{top, tag}}$.

$N_{\text{top, tag}}$ is easy to evaluate using a data-driven method since its corresponding region is a top control region, so we have:

$$N^{\text{estimated}}_{\text{top, tag}} = N_{\text{data, tag}} - N_{\text{non-top,tag}}\,, \tag{5.23}$$

where $N_{\text{data, tag}}$ is the number of data events in the control region and $N_{\text{non-top, tag}}$ the non-top background tagged as top. The top purity of the control region is $\sim 90\%$ and the m_T distribution in this CR in shown Fig. 5.13(a) [49].

For the top tag efficiency $\epsilon_{\text{top, tag}}$, first of all a top control region is defined with two jets and at least one of them being tagged as b-jet. The top tag efficiency

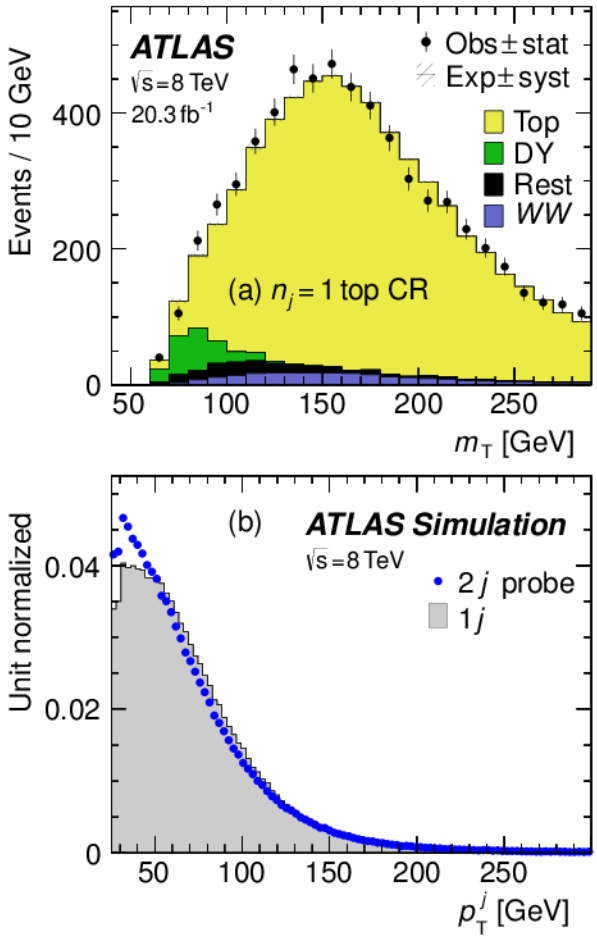


Figure 5.13: Top control region distributions of (a) transverse mass m_T and (b) jet p_T. The m_T plot in (a) scales the top-quark contributions with the normalisation factor calculated here. The p_T^j plot in (b) compares the jet p_T distribution of probing jet in the $n_j = 2$ top control region with that of the jet in the $n_j = 1$ signal region.

can be calculated in this region by a tag-and-probe method and is denoted as $\epsilon^{2j,\mathrm{MC}}_{\text{top, tag}}$. There is an ambiguity in the definition of the tagged jet and the probing jet here. It turns out that a random assignment performs best since in this case the probing jet has similar kinematics as the jet in $n_j = 1$ channel which is shown in Fig. 5.13(b) [49]. All top MC related quantities in the calculation of $\epsilon^{2j,\mathrm{MC}}_{\text{top, tag}}$ could be replaced by the corresponding data-driven values of the non-top subtracted data. Thus we also have a data-driven version of the top tag efficiency calculated from the $n_j = 2$ sample $\epsilon^{2j,\mathrm{estimated}}_{\text{top, tag}}$, on which we can apply the extrapolation factor using

the MC prediction to get the data-driven $n_j = 1$ top tag efficiency:

$$\epsilon^{1j,\text{estimated}}_{\text{top, tag}} = \epsilon^{2j,\text{estimated}}_{\text{top, tag}} \cdot \frac{\epsilon^{1j,\text{MC}}_{\text{top, tag}}}{\epsilon^{2j,\text{MC}}_{\text{top, tag}}} \tag{5.24}$$

$$= \epsilon^{2j,\text{estimated}}_{\text{top, tag}} \cdot \alpha^{\text{MC}}_{\text{eff}} , \tag{5.25}$$

where the MC-based efficiency extrapolation factor $\alpha^{\text{MC}}_{\text{eff}}$ is also defined. The $\epsilon^{1j,\text{MC}}_{\text{top, tag}}$ in the above formula is just the realisation in the $n_j = 1$ channel of the basic Formula 5.21.

Collecting together all the pieces, we have the expression for the top background estimation in $n_j = 1$ channel after the bveto cut:

$$N^{\text{estimated}}_{\text{top, untag}} = \frac{N^{\text{estimated}}_{\text{top, tag}}}{\epsilon^{2j,\text{estimated}}_{\text{top, tag}} \cdot \alpha^{\text{MC}}_{\text{eff}}} \cdot (1 - \epsilon^{2j,\text{estimated}}_{\text{top, tag}} \cdot \alpha^{\text{MC}}_{\text{eff}}) . \tag{5.26}$$

The top MC related systematics lie in the extrapolation factor $\alpha^{\text{MC}}_{\text{eff}}$. And the statistical uncertainty is mainly from the data in the control regions. There are also other small systematics from the non-top processes subtraction.

By dividing $N^{\text{estimated}}_{\text{top, untag}}$ with $N^{\text{MC}}_{\text{top, untag}}$, the normalisation factor for top background in the $n_j = 1$ signal region after the bveto cut can be calculated to be: 1.06 ± 0.03(stat.). To estimate the top background after all selections of the signal region as well as in the WW $n_j = 1$ control region, $N^{\text{estimated}}_{\text{top, untag}}$ is extrapolated to these regions by MC predicted extrapolation factors:

$$\alpha_{\text{untag}\to\text{SR}} = \frac{N_{\text{top, untag}}}{N_{\text{top, SR}}} , \quad \alpha_{\text{untag}\to WW\text{ CR}} = \frac{N_{\text{top, untag}}}{N_{\text{top, }WW\text{ CR}}} . \tag{5.27}$$

Top in the VBF $n_j \geq 2$ channel

The top quark background is the leading background in the $n_j \geq 2$ channel due to the nature of multiple jets (two high-p_T b-jets + possible radiation jets) in its final state. Even there is a bveto cut in the signal region selection, it is still the dominant background. These survived top events mainly have a light-quark jet from ISR and b-jet untagged due to the inefficiency of the tagger. So the control region is defined by requiring $n_j = 1$ rather than $n_j \neq 0$ (although the latter brings additional statistics for the CR) since in this case the kinematics of the two leading jets are closer to that of the SR. Besides, the background estimation is carried out using all lepton flavour channels, which is different from the previous cases, since

the Drell-Yan contamination is no longer large due to high jet multiplicity nature of this channel.

The methodology of the top estimation is very simple which is based on Formula 5.8, The top control region for the BDT-based analysis is defined after the common preselection and the channel-dependent selection ($n_j \geq 2$):

- $n_{b\text{-jet}} = 1$: as explained, this definition is to mimic the jet flavour composition of the signal region,
- CJV + OLV + $Z\tau\tau$ veto.

The m_{jj} and output BDT distributions are compared between data and MC in Fig. 5.14 [49] for the top CR defined above.

Due to the large difference of top event kinematics over each output BDT bin, the normalisation factors are calculated separately in each bin, with the two highest score bins merged because of lacking of statistics. So there are in total two normalisation factors: 1.58 ± 0.15(stat.) for the lower score bind and 0.95 ± 0.31(stat.) for the higher one.

As for the cut-based BDT analysis, the normalisation factors are calculated in each signal region cut stage. And the only difference of cuts between SRs and CRs is just the altering of the bveto to $n_{b\text{-jet}} = 1$. After splitting the final SR (corresponds to the last signal region cut) into two m_{jj} bins, the NFs in each of them are 0.94 and 1.48, consistent with the BDT-based analysis.

Top in the ggF $n_j \geq 2$ channel

The top background is the leading background in this channel even after the bveto cut, which is similar as the above VBF case. Moreover, the same method is also used. However, the control region definition is different: a $m_{\ell\ell} > 80\,\text{GeV}$ cut, with the $m_{\ell\ell}$ distribution shown in Fig. 5.15 [49], is used to make the top CR orthogonal to the SR rather than a $n_{b\text{-jet}}$ cut. This is motivated by the reduced systematics of the former definition and the similarity $t\bar{t}$ versus Wt composition between the SR and the former CR. The resulting control region is of $\sim 70\%$ purity and the normalisation factor is: 1.05 ± 0.03(stat.).

Top in 7 TeV

The top background estimation in the 7 TeV analysis follows exactly the same procedures described above. Except that in the $n_j = 1$ channel, a simpler method

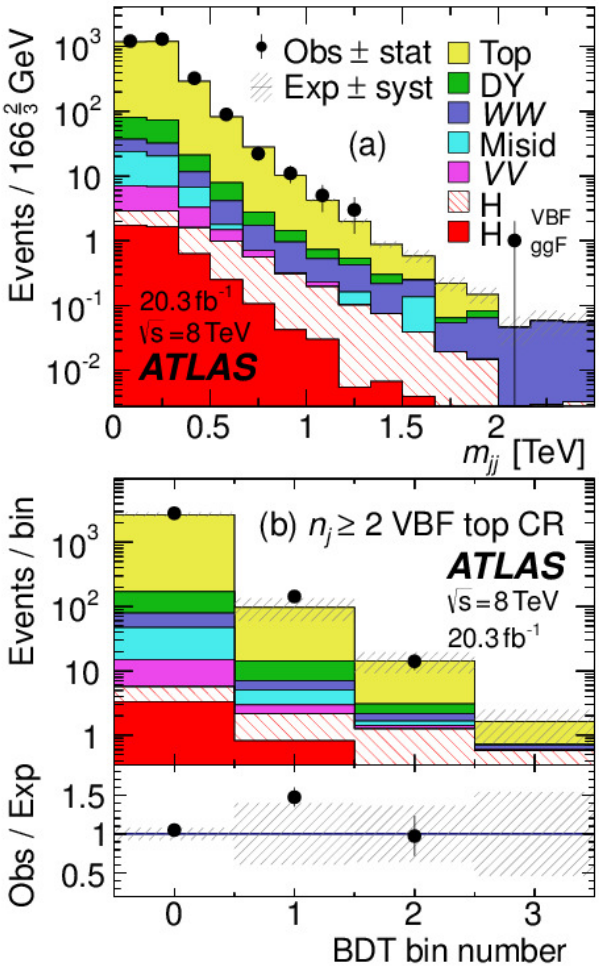


Figure 5.14: Top control region distributions in the VBF $n_j \geq 2$ channel for the BDT-based analysis: (a) m_{jj} and (b) BDT output. For the plot in (b) the shaded band in the ratio shows the uncertainty on the normalisation of each bin. No events are observed in bin 3.

is used with the control region defined by inverting the bveto cut and the extrapolation factor is purely MC based.

5.4.3 W+jets and QCD

As mentioned earlier, the W+jets and QCD shown in the signal region are due to one or two jets in their final state being reconstructed as leptons so that they pass the lepton number requirement. Although the probability is small, they become non-negligible due to their huge production cross sections. This misidentification rate is not well modelled in MC samples. So a data-driven method called *Fake Factor Method* is developed for the estimation of these backgrounds and applied to

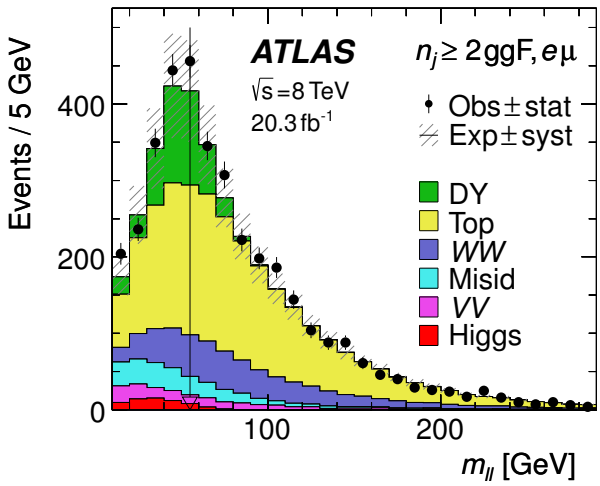


Figure 5.15: Distribution of dilepton invariant mass for the ggF $n_j \geq 2$ channel. The plot is made after requiring all selections up to $m_{\ell\ell}$.

all lepton flavour and jet multiplicity channels.

W+jets

The method works in a similar way as Formula 5.8, with the extrapolation factor α replaced by the so-called fake factor. There is a substantial difference between these two similar-looking quantities: the calculation of the fake factor is more flexible than that of α. In Formula 5.8, α is calculated by MC predictions and the control region is the same region where the extrapolation will be applied to, while for the fake factor, both conditions can be changed.

Let's start by defining the W+jets control region. The common rule for choosing a control region still works here: inverting one signal region selection. In this case, it is one of the good lepton, called "id" lepton here, be changed to the "anti-id" lepton, which is defined by failing the good lepton criteria and satisfying a loosened lepton selection [55]. The anti-id lepton has a large fake rate by definition or in other words is jet-enriched. By replacing requirement of the two id leptons in the signal region selection to one id lepton and one anti-id lepton, the W+jets control region is constructed whose number of events can be expressed as:

$$N_{\text{id+anti-id}} = N^{W+\text{jets}}_{\text{id+anti-id}} + N^{\text{QCD}}_{\text{id+anti-id}} + N^{\text{EW}}_{\text{id+anti-id}} \,. \tag{5.28}$$

In the expression the contributions from W+jets, QCD and other processes (mainly electroweak processes) are separated. So the number of W+jets events can be

estimated, in the control region to be:

$$N_{\text{id+anti-id}}^{W\text{+jets, est}} = N_{\text{id+anti-id}}^{\text{data}} - N_{\text{id+anti-id}}^{\text{QCD}} - N_{\text{id+anti-id}}^{\text{EW, MC}}\,, \tag{5.29}$$

where the QCD contribution is subtracted using a data-driven estimation and others using MC predictions. Besides, the signal region can be written in a similar style:

$$N_{\text{id+id}} = N_{\text{id+id}}^{W\text{+jets}} + N_{\text{id+id}}^{\text{QCD}} + N_{\text{id+id}}^{\text{EW}}\,. \tag{5.30}$$

The fake factor plays the role of extrapolation factor and is simply defined as:

$$f_l = \frac{N_{j,\text{id}}}{N_{j,\text{indi-id}}},\ (l = e \text{ or } \mu)\,, \tag{5.31}$$

where $N_{j,\text{id}}$ ($N_{j,\text{anti-id}}$) is the number of jets fulfilling the id (anti-id) lepton selection. It is measured in a control sample of jet production associated with a Z boson with subsequent decay into ee or $\mu\mu$ (in the 7 TeV analysis, it is instead measured in dijet sample). The Z boson sample is used to tag the control region which could be done using a Z-mass window cut. Besides, there are other cuts to suppress the contamination of other processes falling into the mass window. The jets are the probing object, which may be reconstructed and selected as id lepton or anti-id lepton, whose p_T distributions are shown in Fig. 5.16 [49]. The fake factor in each p_T bin is just the ratio of the events in the upper plots over those in the lower plots with the contamination from other EW processes subtracted using MC. The final measured fake factors are in the order of 0.01 (0.1) for electron (muon) with a p_T and η dependence.

So that the predicted W+jets in the signal region can be expressed as:

$$N_{\text{id+id}}^{W\text{+jets, est}} = N_{\text{id+anti-id}}^{W\text{+jets, est}} \cdot f_l. \tag{5.32}$$

The estimated W+jets in this way suffer from a large systematic uncertainty which is mainly from the uncertainty of the extrapolation factor f_l. The error of f_l comes from the different jet flavour compositions between Z+jets sample which is used to extract the fake factor and the W+jets sample on which the fake factor to be applied. Besides, the low statistics of the probing jet sample and the subtraction of the EW contamination also contributes significantly. In total, the relative systematic errors range between 29% to 61% for jet faking as electron and 25% to 46% for muon.

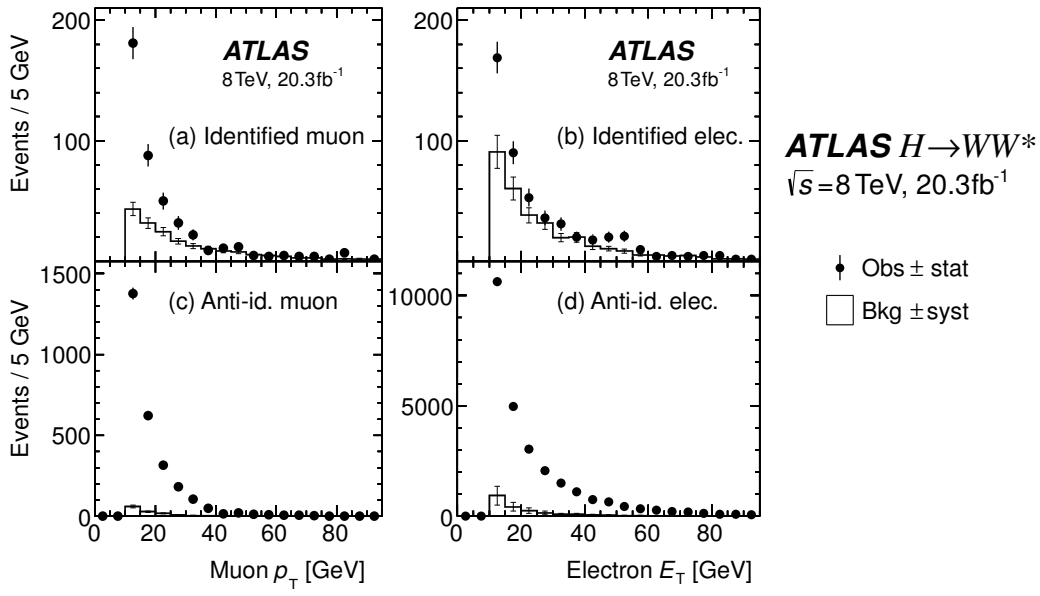


Figure 5.16: Distributions of p_T for jets misidentified as id-lepton or anti-id lepton in the Z_{jet} control sample: (a) identified muon, (b) identified electron, (c) anti-identified muon, and (d) anti-identified electron. The dots represent the data (Obs). The histograms are the background MC estimates (Bkg) of the sum of electroweak processes other than the associated production of a Z boson and jets.

It is also required to estimate the W+jets background in the same sign (SS) region, which is defined by just inverting the "opposite sign" cut of the signal selection. The above fake factor cannot be used in this case. Because the W+jets process is not expected to produce equal numbers of SS and OS candidates. In particular, associated production processes such $W + c$, where the second lepton comes from the semileptonic decay of a charmed hadron, produce dominantly OS candidates. So the above method is applied in the SS region resulting in another fake factor.

QCD

The QCD background is estimated similarly as for the W+jets with the control region defined by replacing the requirement of the two id leptons in the signal region selection to two anti-id leptons. So that the number of QCD events in the CR is expressed as:

$$N^{\text{QCD, est}}_{\text{anti-id+anti-id}} = N^{\text{data}}_{\text{anti-id+anti-id}} - N^{W\text{+jets, MC}}_{\text{anti-id+anti-id}} - N^{\text{EW, MC}}_{\text{anti-id+anti-id}} , \tag{5.33}$$

where both W+jets and others EW contamination are subtracted using MC predictions.

The fake factor for the QCD background is measured in a dijet sample with one fake lepton, rather than using the one for W+jets, since the dijet sample is more close to the QCD background in terms of jet flavour composition. With the fake factor, the QCD contamination in the W+jets control sample is:

$$N^{\text{QCD, est}}_{\text{id+anti-id}} = N^{\text{QCD, est}}_{\text{anti-id+anti-id}} \cdot 2 \cdot f'_l, \tag{5.34}$$

the factor 2 comes from the fact that either jet may be misidentified. And in this case, f'_l is the corrected fake factor for possible sample dependence since we are applying the fake factor from the dijet sample containing an anti-id lepton to the dijet sample with two anti-id leptons. This fake factor is applied twice because of the two misidentified leptons that bring the QCD into signal region. As for the estimation of QCD in the signal region, it is:

$$N^{\text{QCD, est}}_{\text{id+id}} = N^{\text{QCD, est}}_{\text{anti-id+anti-id}} \cdot f'_l \cdot f''_l, \tag{5.35}$$

f''_l is the corrected fake factor for the possible sample dependence since the original fake factor is now applied to the dijet sample with an id lepton and an anti-id lepton. Besides, same as W+jets, the above procedures are also applied in the same sign region to derive a SS fake factor for the estimation of the QCD contribution in the SS region.

Summary

In practice, the above data-driven methods are applied before the actual event selection. This means that, rather than defining the control regions and applying the extrapolation factors after the event selection began like those procedures for the WW and top, the W+jets and QCD control samples are reweighted by the fake factors to form the data-driven W+jets and QCD samples before cuts are applied. These samples play the role of the MC samples and can be used directly. It is also due to this special treatment, the tasks of W+jets and QCD estimation in each sub-analyses are done at one stroke.

5.4.4 Non-WW dibosons

The non-WW diboson backgrounds ($W\gamma$, $W\gamma^{(*)}$, WZ, ZZ, and $Z\gamma$) are amount to $\sim 10\%$ of the total backgrounds in the $n_j \leq 1$ analyses, but very limited in the $n_j \geq 2$ channels (so MC prediction is used). Among the five processes, the $W\gamma$, $W\gamma^{(*)}$ and WZ processes dominate. They mimic the signal signature by the W leptonic decay which provides a prompt lepton and a neutrino (missing energy) and the γ, γ^*, or Z decaying into a lepton pair ($ee/\mu\mu$, in case of γ only electron pair) with only one of them being identified. As for the ZZ, there are two possibilities: if both Z bosons decay into lepton pairs, it falls into the signal region when two of the four leptons are unidentified, which is vary rare and if one Z boson decays into lepton pair and the other two neutrino pairs, it has the same signature as the signal.

For the three dominant diboson processes ($W\gamma$, $W\gamma^{(*)}$, and WZ) in the $n_j \leq 1$ analyses, it is found that they are equally likely to produce a second lepton of either charge with respect to the charge of the lepton from W decay. If the charges of the two leptons are opposite, they enter into the opposite sign (OS) region, which is the signal region, else they go to the same sign (SS) region. And the normalisation and kinematics should be identical to a very good approximation between the OS and SS regions. But this only suits for the $e\mu/\mu e$ channel since in the $ee/\mu\mu$ channel, the population in the SS and OS regions are not identical which will bring additional uncertainties when doing the extrapolation. Thus the same sign control region (SSCR) can be defined to normalise these backgrounds both in $n_j = 0$ and $n_j = 1$ analysis by just inverting the "opposite sign leptons" requirement of the signal event selections. These non-WW dibosons takes up to $\sim 60\%$ of the SSCR events with the rest being mainly W+jets and QCD. The MC modelling in the SSCR is checked by looking at the m_T and p_T^{sublead} distributions as shown in Fig. 5.17 [49].

Normalisation factors can be calculated as in Formula 5.8, which are 0.92 ± 0.07(stat.) for the $n_j = 0$ channel and 0.96 ± 0.12(stat.) for the $n_j = 1$ channel. Besides, due to the very small portion of the rest two non-WW diboson processes (ZZ and $Z\gamma$), they are normalised together with the three dominating ones. However, the relative contributions of the individual processes are still estimated using MC samples.

For those channels where the non-WW dibosons are not constrained by control regions, there exist validation regions to check the normalisation and shape mod-

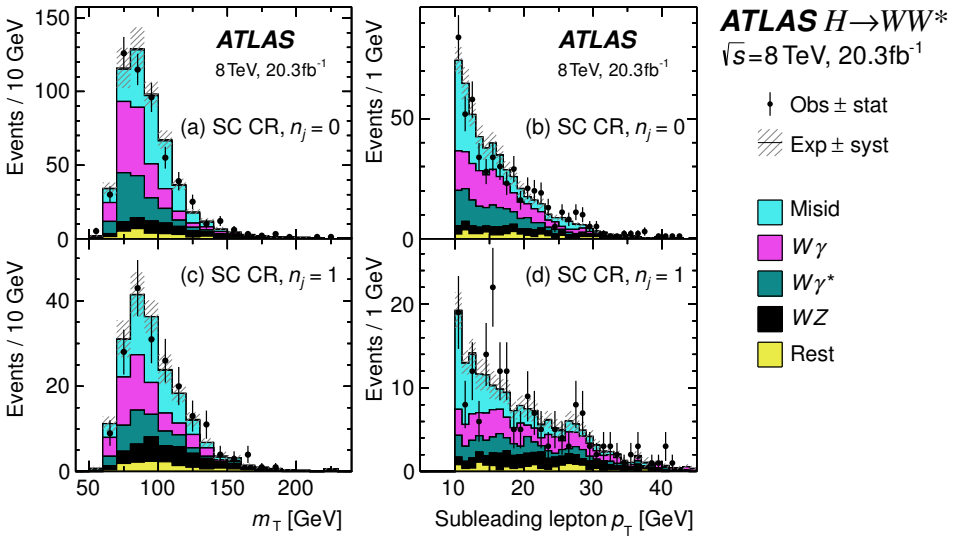


Figure 5.17: Same sign control region distributions: (a) transverse mass in the $n_j = 0$ category, (b) sub-leading lepton p_T in the $n_j = 0$ category, (c) transverse mass in the $n_j = 1$ category, and (d) sub-leading lepton p_T in the $n_j = 1$ category. "Rest" consists of contributions not listed in the legend.

elling of $W\gamma$, $W\gamma^*$ and $Z\gamma$, among which the $Z\gamma$ validation region is used to derive photon conversion uncertainty. Similarly, the non-WW diboson estimations in the 7 TeV analysis are all from MC predictions, which are checked in the validation regions.

5.4.5 Drell-Yan

The Drell-Yan processes contaminate the signal region in two ways. In the case of the Z/γ^* decaying into $\tau\tau$ and both τ undergo leptonic decay, there will be neutrinos resulting in missing energy in the final state. Together with the two leptons from the τ decays, the events possess similar signature as the signal. While in the case of Z/γ^* decaying into $ee/\mu\mu$, events pass the E_T^{miss} cut due to the limited resolution of the detector at high pile-up condition and the neutrinos from b-hadron or c-hadron decays. The $\tau\tau$ mode is mainly a problem of the $e\mu/\mu e$ analysis while the other mode is relevant to $ee/\mu\mu$.

$Z/\gamma^* \to \tau\tau$

The estimation of $Z/\gamma^* \to \tau\tau$ in the ggF dedicated channels is done in the $e\mu/\mu e$ flavour and applied to all lepton flavour channels (except for ggF $n_j \geq 2$

since only $e\mu/\mu e$ is considered). The simple method as described by Formula 5.8 is used.

For the $n_j = 0$ analysis, the control region is defined by cuts after the common preselection and jet veto:

- $m_{\ell\ell} < 80\,\text{GeV}$: the dilepton invariant mass is constrained by the mass of its mother particle Z/γ^*,
- $\Delta\phi_{\ell\ell} > 2.8$: the two leptons from Z/γ^* decay tend to be back-to-back.

The control region has a 91% purity of $Z/\gamma^* \to \tau\tau$ and a 0.5% contamination from $Z/\gamma^* \to ee/\mu\mu$. The resulting normalisation factor is $1.00 \pm 0.02(\text{stat.})$.

For the $n_j = 1$ analysis, the control region is defined after the $M_{T,\text{max}}^W$ cut of the signal region selection by:

- $m_{\ell\ell} < 80\,\text{GeV}$: similar as the $n_j = 0$ control region,
- $m_{\tau\tau} > m_Z - 25\,\text{GeV}$: to make it orthogonal to the $n_j = 1$ signal region.

The $Z/\gamma^* \to \tau\tau$ purity in the control region is 80% and its normalisation factor is evaluated to be: $1.05 \pm 0.04(\text{stat.})$.

As shown in Fig. 5.10, the $Z/\gamma^* \to \tau\tau$ has sizeable contribution in the WW control regions of the $n_j \leq 1$ channels, which is also constrained by the control regions defined above.

For the ggF $n_j \geq 2$ analysis, the control region is defined after the bveto cut of the signal region selection by:

- $\Delta\phi_{\ell\ell} > 2.8$: similar as the $n_j = 0$ control region,
- $m_{\ell\ell} < 70\,\text{GeV}$: similar as the $n_j = 0$ control region,
- Fail either CJV or OLV: to be orthogonal to the VBF $Z/\gamma^* \to \tau\tau$ control region.

The purity of $Z/\gamma^* \to \tau\tau$ in the control region is 74% with a normalisation factor of $1.00 \pm 0.09(\text{stat.})$.

For the estimation of $Z/\gamma^* \to \tau\tau$ in the BDT-based VBF analysis, the same method as in the ggF analysis is used, except that all four lepton flavour channels are included in the control region, which is defined after the common preselection and $n_j \geq 2$ requirement:

- $m_{\ell\ell} < 80(75)\,\text{GeV}$ for $e\mu/\mu e$ ($ee/\mu\mu$): similar as the $n_j = 0$ control region,
- $|m_{\tau\tau} - m_Z| < 25\,\text{GeV}$: similar as the $n_j = 1$ control region,
- BDT score > -0.48: remove the 1st background-enriched BDT bin to make the control region closer to the signal region.

The normalisation factor is calculated to be $0.9 \pm 0.3(\text{stat.})$.

$Z/\gamma^* \to ee/\mu\mu$ in the $n_j \leq 1$ analyes

The method used for the estimation of $Z/\gamma^* \to ee/\mu\mu$ events in the $n_j \leq 1$ channels, called Pacman method [58–60], is a bit complicated than the above cases. It is based on the feature of one of the variable used in the signal region selection, which is the f_{recoil} or $f_{\text{recoil}}^{\text{extended}}$. Take the f_{recoil} as an example, whose distribution is shown in Fig. 5.3(d), it can be seen that the shape of the Drell-Yan process in $ee/\mu\mu$ (dominated by $Z/\gamma^* \to ee/\mu\mu$ and in green histogram) has a difference against the rest processes (denoted as non-DY here).

The Pacman method is illustrated in Fig. 5.18 [55]. The key is a template fit of the f_{recoil} (or $f_{\text{recoil}}^{\text{extended}}$) data distribution in the final signal region, which is the red box. The template used in the fit is from two other templates:

$$T_{\text{SR}} = N_{\text{non-DY}} \cdot T_{\text{non-DY}} + N_{\text{DY}} \cdot T_{\text{DY}}\,, \tag{5.36}$$

where T_{DY} and $T_{\text{non-DY}}$ are the templates for the Drell-Yan and non-Drell-Yan processes. N_{DY} and $N_{\text{non-DY}}$ are free parameters of the signal region template T_{SR}. The non-DY template $T_{\text{non-DY}}$ is extracted from the $e\mu/\mu e$ channel $T_{\text{data}}^{e\mu/\mu e}$ which is a $\sim 100\%$ non-DY control region: The DY template T_{DY} is extracted from the Z peak region which is defined by inverting the Z mass window cut of the signal region selection:

$$T_{\text{DY}} = T_{\text{data}}^{Z\text{ peak}} - N_{\text{non-DY}}^{Z\text{ peak}} \cdot T_{\text{non-DY}}^{Z\text{ peak}}\,. \tag{5.37}$$

Since the Z peak region is only 50% pure, the non-DY contamination to the template is subtracted. The subtraction makes use of the MC predicted non-DY event yield so that $N_{\text{non-DY}}^{Z\text{ peak}} = N_{\text{non-DY}}^{Z\text{ peak MC}}$ and the data-driven non-DY shape which is drawn again from the corresponding $e\mu/\mu e$ phase space so that $T_{\text{non-DY}}^{Z\text{ peak}} = T_{\text{data}}^{Z\text{ peak }e\mu/\mu e}$ (which is normalised to unity).

Substitute all the above formulas into Formula 5.36, we have the estimated

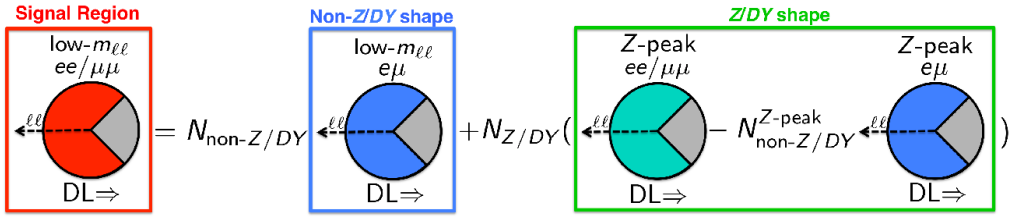


Figure 5.18: Scheme representing the Z/DY estimate procedure (the Pacman method).

signal region template:

$$T_{\text{SR}} = N_{\text{non-DY}} \cdot T_{\text{data}}^{e\mu/\mu e} + N_{\text{DY}} \cdot (T_{\text{data}}^{Z\text{ peak}} - N_{\text{non-DY}}^{Z\text{ peak MC}} \cdot T_{\text{data}}^{Z\text{ peak }e\mu/\mu e}), \tag{5.38}$$

which has a one-to-one correspondence to elements in Fig. 5.18.

From the fit, both Drell-Yan and non-Drell-Yan in the signal region can be estimated. In practice, for simplicity the template is a 2-bin distribution with the boundary set to be the cut value of f_{recoil} (or $f_{\text{recoil}}^{\text{extended}}$). In this case, the templates degrade into cut efficiencies: $T_{\text{non-DY}} \to \epsilon_{\text{non-DY}}$, $T_{\text{non-DY}}^{Z\text{ peak}} \to \epsilon_{\text{non-DY}}^{Z\text{ peak}}$, and $T_{\text{DY}} \to \epsilon_{\text{DY}}$.

$Z/\gamma^* \to ee/\mu\mu$ in the BDT-based VBF analysis

The $Z/\gamma^* \to ee/\mu\mu$ background in the VBF $ee/\mu\mu$ channel is estimated using an ABCD method, which is illustrated in Table 5.12 [55]. The "CRE" means a low E_T^{miss} control region while the "CRM" a Z mass window control region.

The estimated number of DY events ($Z/\gamma^* \to ee/\mu\mu$) in the signal region can be expressed as:

$$N_{DY}^{\text{est},A} = N_{\text{DY}}^{\text{est},B} \cdot \frac{N_{\text{DY}}^{\text{est},C}}{N_{\text{DY}}^{\text{est},D}} \cdot f_{\text{correction}} \tag{5.39}$$

where $N_{\text{DY}}^{\text{est},B}$, $N_{\text{DY}}^{\text{est},C}$, and $N_{\text{DY}}^{\text{est},D}$ are the data-driven estimation of DY events in each region. $\frac{N_{\text{DY}}^{\text{est},C}}{N_{\text{DY}}^{\text{est},D}}$ plays the role of extrapolation factor. $f_{\text{correction}}$ is defined as:

$$f_{\text{correction}} = \frac{N_{\text{DY}}^{\text{MC},A}/N_{\text{DY}}^{\text{MC},B}}{N_{\text{DY}}^{\text{MC},C}/N_{\text{DY}}^{\text{MC},D}}, \tag{5.40}$$

which is to correct the possible bias of applying the extrapolation factor calculated from regions C and D to regions A and B.

The normalisation factor is calculated separately for the second BDT bin and the 3rd+4th BDT bins, with the values to be: 1.03 ± 0.15(stat.) and $0.89 \pm$

Table 5.12: Summary of the regions used for the Z_{DY} estimation technique used in the SF channel of the VBF BDT analysis.

Region A (SR) $E_T^{\mathrm{miss,calo}} > 45\,\mathrm{GeV}$ $m_{\ell\ell} < 75\,\mathrm{GeV}$	**Region C (Z CRM)** $E_T^{\mathrm{miss,calo}} > 45\,\mathrm{GeV}$ $\lvert m_{\ell\ell} - m_Z \rvert < 15\,\mathrm{GeV}$
Region B (Z CRE) $25\,\mathrm{GeV} < E_T^{\mathrm{miss,calo}} < 45\,\mathrm{GeV}$ $m_{\ell\ell} < 75\,\mathrm{GeV}$	**Region D (Z CRM)** $25\,\mathrm{GeV} < E_T^{\mathrm{miss,calo}} < 45\,\mathrm{GeV}$ $\lvert m_{\ell\ell} - m_Z \rvert < 15\,\mathrm{GeV}$

0.28(stat.).

5.4.6 Summary

All the control regions used in the above background estimation procedures are summarised in Table 5.13 [49]. For those special methods, e.g. the Pacman method, the control regions are not shown in the table. The composition of each processes is detailed as well as the purity of the control region.

In Table 5.14 [49], all normalisation factors derived from the above listed control regions are shown, with only statistical uncertainties. In most cases, the NFs are consistent with unity and for those not, if taking into account the systematic uncertainties, will recover the consistency. However, there is an exception which is the WW $n_j = 0$ normalisation factor. Even after considering systematic, it is $1.22 \pm 0.03(\mathrm{stat.}) \pm 0.10(\mathrm{sys.})$ showing a $2\,\sigma$ deviation.

5.5 Theoretical systematics

The systematics appear in almost every part of the analysis and determine the quality of the analysis: the larger they are, the less significant conclusion can be drawn. Take the WW $n_j = 0$ normalisation factor as an example in the previous section, if no systematic is estimated, the NF is significantly deviating from unity which strongly indicates that the model is wrong. However, taking into account the systematics, the deviation lowered to $2\,\sigma$, not enough to establish an evidence. Thus it is important to have a complete and accurate assessment on these uncertainties.

Table 5.13: Control region event yields for 8 TeV data. All of the background processes are normalised with the corresponding normalisation factors given in Table 5.14 or with the data-derived methods as described in the text; each row shows the composition of one CR. The N_{sig} column includes the contributions from all signal production processes. For the VBF channel, the values for the bins in output BDT are given. The entries that correspond to the target process for the CR are given in bold; this quantity corresponds to N_{bold} considered in the last column for the purity of the sample (in %). The uncertainties on N_{bkg} are due to sample size.

Control regions	Summary			Composition of N_{bkg}						Purity
	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{top}	N_{misid}	N_{VV}	N_{DY} $N_{ee/\mu\mu}$	N_{DY} $N_{\tau\tau}$	N_{bold}/N_{bkg} (%)
$n_j = 0$										
CR for WW	2713	2680±9	28	**1950**	335	184	97	8.7	106	73
CR for top quarks	76013	75730±50	618	8120	**56210**	2730	1330	138	7200	74
CR for VV	533	531±8	2.2	2.5	1.1	180	**327**	19	2.7	62
CR for $Z/\gamma^* \to \tau\tau$	4557	4530±30	23	117	16.5	239	33	28	**4100**	91
$n_j = 1$										
CR for WW	2647	2640±12	4.3	**1148**	1114	165	127	17	81	43
CR for top quarks	6722	6680±12	17	244	**6070**	102	50	6	204	91
CR for VV	194	192±4	1.9	1	3.1	65	**117**	4.7	0.8	61
CR for $Z/\gamma^* \to \tau\tau$	1540	1520±14	18	100	75	84	27	7	**1220**	80
$n_j \geq 2$ ggF										
CR for top quarks	2664	2660±10	4.9	561	**1821**	129	101	10	44	68
CR for $Z/\gamma^* \to \tau\tau$	266	263±6	2.6	13	34	18	4.1	0.1	**194**	74
$n_j \geq 2$ VBF										
CR for top quarks, bin 1	143	142±2	2.1	1.9	**130**	2.1	0.8	6.3	1.1	92
CR for top quarks, bin 2–3	14	14.3±0.5	1.8	0.6	**11.6**	0.2	0.2	0.9	0.2	81
CR for $Z/\gamma^* \to \tau\tau$	24	20.7±0.9	2.4	0.9	1.2	0.6	0.2	0.8	**17**	82

Table 5.14: Summary of control region normalisation factors. Those that use MC-based normalisation are marked with a dash. For the VBF $n_j \geq 2$ events, the values in bins of output BDT are given for top quarks; a combined value is given for $Z/\gamma^* \to \tau\tau$. The uncertainties are due to the sample size of the corresponding control regions.

Channels	WW	Top	VV	$Z/\gamma^* \to \tau\tau$
8 TeV sample				
$n_j = 0$	1.22 ± 0.03	1.08 ± 0.02	0.92 ± 0.07	1.00 ± 0.02
$n_j = 1$	1.05 ± 0.05	1.06 ± 0.03	0.96 ± 0.12	1.05 ± 0.04
$n_j \geq 2$, ggF	-	1.05 ± 0.03	-	1.00 ± 0.09
$n_j \geq 2$, VBF bin 1	-	1.58 ± 0.15	-	0.90 ± 0.30
$n_j \geq 2$, VBF bins 2–3	-	0.95 ± 0.31	-	0.90 ± 0.30
7 TeV sample				
$n_j = 0$	1.09 ± 0.08	1.12 ± 0.06	-	0.89 ± 0.04
$n_j = 1$	0.98 ± 0.12	0.99 ± 0.04	-	1.10 ± 0.09
$n_j \geq 2$, VBF bins 1–3	-	0.82 ± 0.29	-	1.52 ± 0.91

The systematic sources are divided into two categories: theoretical uncertainty and experimental uncertainty. The theoretical part could come from limited pre-

cision of the perturbative calculation for the hard process, limited understanding of the non-perturbative QCD which is important to understand the underlying events or the parton radiation, the uncertainty in the PDF measurement etc. The experimental part could come from the mis-modelling of the detector response, the mis-modelling of the pile-up, the inefficiency of the reconstruction softwares etc.

The theoretical uncertainty of each process will be introduced in this section while the experimental uncertainty will be discussed in the next section.

Theoretical uncertainty exists in the MC predicted normalisation and distribution. For the signal processes, it lies mostly on the cross section calculation both in terms of normalisation and distribution. For the backgrounds, since there are always control regions available (except for the less important backgrounds like DPI), only the uncertainty of the extrapolation factor, which is in fact a distribution, needs to be evaluated.

The theoretical uncertainty comes in every step of the sample production: from the perturbative choosing the QCD scales of the calculation to the sampling the parton PDFs then to the hard process and the dressing of soft processes. If there are other processes with the same final state as the signal process, quantum interference effect should be considered. Thus in general uncertainties on the following aspects are considered: the QCD scale (factorisation scale and renormalisation scale), the PDF, the parton shower and underlying events (UE/PS) modelling, and their match with ME generator.

The QCD scale uncertainty is to cover the missing high order corrections not applied (available) and is usually evaluated by varying up/down the factorisation scale and renormalisation scale independently by a factor of 2 and choosing the maximum as the uncertainty [61].

The PDF uncertainty includes the uncertainty from different parametrisation and models used by different groups which is evaluated by comparing two different PDF sets (like CT10 versus MSTW), and the uncertainty from input data to the PDF measurement which is usually estimated using error eigenvectors provided by the PDF set (like CT10).

The same ME generator is interfaced with different UE/PS models to evaluate the UE/PS modelling uncertainty and different ME generators are interfaced to the same UE/PS to probe the matching uncertainty.

In the following, the theoretical uncertainties will be introduced in detail from the signal process to each background processes.

5.5.1 Signal

The signal event yields in the selected signal regions determine directly the final conclusion of the analysis. Thus their uncertainties are very important. Moreover, there is no control regions to rely on which could help to cancel theoretical systematic effects. So the signal prediction is always carried out to highest possible orders.

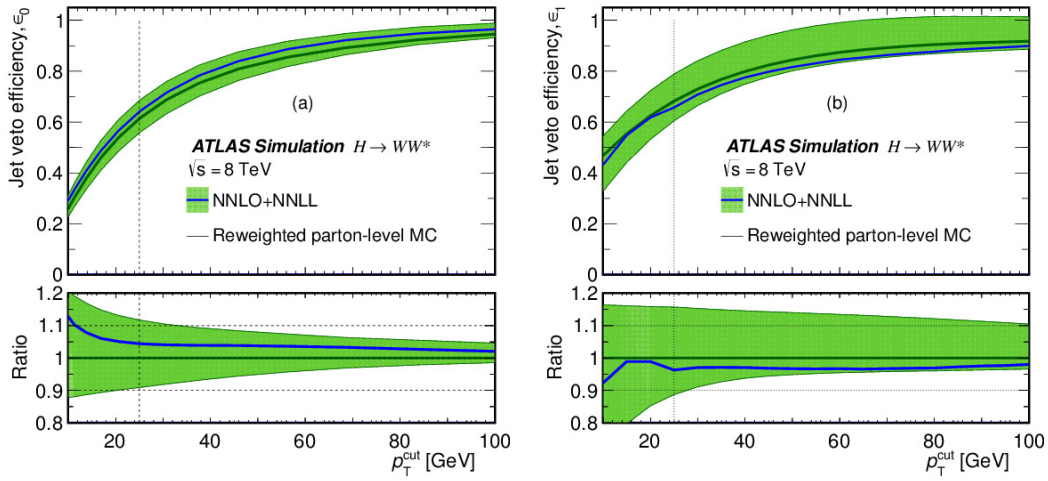
The theoretical uncertainties lie in both the Higgs production (cross section) and Higgs decay (branching ratio). The branching ratio is calculated to be 22% $\pm$ 4.2% for $m_H = 125.36\,\text{GeV}$ using PROPHECY4F [62]. While the cross section treatments are described below.

ggF

The lowest order Feynman diagram of the ggF process is already shown in Fig. 3.7. To compute the higher order QCD corrections to the Higgs production cross section, radiations from the initial state gluons and the top quark loop are included.

For the inclusive production cross section, it is computed to next-to-next-to-leading order (NNLO) [63] using an effective point-like ggH coupling by assuming top mass $m_t \to \infty$. Then the finite top mass effects are corrected from next-to-leading order (NLO) calculation and found to be a few percent [64]. Resummation of the soft QCD radiation has been performed to the next-to-leading log (NLL) for finite top and bottom masses and the next-to-next-to-leading log (NNLL) [65] in infinite top mass assumption. As for electroweak corrections, it is computed to NLO [66] using the complete factorisation approximation [67]. All in all, the inclusive ggF production cross section is calculated to be: 19.15 pb for a Higgs mass of 125.36 GeV with a 10% uncertainty which receiving contributions from QCD scale variation (7.5%) and PDF (7.2%) [49].

As for the exclusive cross section, the most important one is jet multiplicity distribution, since the analysis is divided into sub-channels according to it. This distribution is modelled by POWHEG+Pythia8 which is in NLO precision with the Higgs p_T distribution reweighted to NNLO+NNLL prediction of HRES[68] for $n_j \leq 1$ and to dedicated NLO prediction [69] for $n_j \geq 2$. The uncertainty of the jet multiplicity distribution is evaluated by a jet-veto-efficiency (JVE) method [68, 70] which factorises exclusive cross sections in the three jet bins ($n_j = 0$, $n_j = 1$, and $n_j \geq 2$) into an inclusive cross section and two jet veto efficiencies (one on



(a) $n_j = 0$ (b) $n_j = 1$

Figure 5.19: Efficiencies of the veto of the (a) first jet and (b) second jet in inclusive ggF production of the Higgs boson, as a function of the jet p_T threshold used in the jet selection.

first jet and the other on second jet). The inclusive cross section is already mentioned. The two efficiencies are calculated using JETVHETO [71] and presented in Fig. 5.19 [49] together with their uncertainties as a function of jet p_T threshold. The final uncertainties for the three jet bins are 15%, 27%, and 34% for $n_j = 0$, $n_j = 1$, and $n_j \geq 2$.

Not all phase space of the signals is to be selected by the event selection, thus there are the so-called acceptance uncertainties, which is in fact from the uncertainties of some variable distributions on which the cuts are applied. The four typical systematic sources mentioned previously are considered and all are small (a few to several percent) compared with the above inclusive/exclusive cross section uncertainties. The QCD scale uncertainty is computed as usual. The PDF uncertainties emerges from CT10 versus MSTW2008NLO and CT10 error eigenvectors. The ME-UE/PS matching uncertainty is evaluated by comparing POWHEG+Herwig with aMC@NLO+Herwig, and the UE/PS modelling by comparing POWHEG+Herwig with POWHEG+Pythia8.

As described in the VBF event selection section, the ggF $n_j \geq 2$ signal is considered as background there. The jet bin related uncertainty needs a separate estimation since there is a central jet veto cut in the cut list which makes use of a third jet. This uncertainty is evaluated using Stewart-Tackman (ST) [72] method to be 29%. The acceptance uncertainties similar as those in the previous paragraph

Table 5.15: Signal uncertainties (in %) due to the modelling of the ggF and VBF processes. For the $n_j \leq 1$ categories the uncertainties are shown for events with same-flavour leptons; For the $n_j \geq 2$ VBF category the uncertainties are shown for the most sensitive bin of BDT output (bin 3).

Uncertainty source	$n_j = 0$	$n_j = 1$	$n_j \geq 2$ ggF	$n_j \geq 2$ VBF
ggF				
Total cross section	10	10	10	7.2
Jet binning or veto	11	25	33	29
Acceptance				
Scale	1.4	1.9	3.6	48
PDF	3.2	2.8	2.2	-
Generator	2.5	1.4	4.5	-
PS/UE	6.4	2.1	1.7	15
VBF				
Total cross section	2.7	2.7	2.7	2.7
Acceptance				
Scale	-	-	-	3.0
PDF	-	-	-	3.0
Generator	-	-	-	4.2
UE/PS	-	-	-	14

are also estimated and together summarised in Table 5.15 [49].

VBF

The inclusive VBF production cross section is calculated approximately to NNLO in QCD using VBF@NNLO [73], with NLO EW corrections applied by HAWK [74]. The final uncertainty, dominantly from PDF, is 2.7%. Similar as for the ggF signal, acceptance uncertainties are evaluated for the four types. All uncertainties are summarised in Table 5.15.

5.5.2 WW

***WW* in the $n_j \leq 1$ channels**

In the $n_j \leq 1$ channels, the $e\mu/\mu e$ WW control regions are used to constrain WW in both the $e\mu/\mu e$ signal region and the $ee\mu\mu$ signal region. The theoretical uncertainties of the extrapolation factor are estimated. Since the $e\mu/\mu e$ signal region is divided into six small regions, there will be six extrapolation factors to be

Table 5.16: WW theoretical uncertainties (in %) on the extrapolation factor α for $n_j \leq 1$. Total (Tot) is the sum in quadrature of the uncertainties due to the QCD scale, the PDFs, the matching between the hard-process matrix element and the parton shower and underlying event (UE/PS) model (Gen), and the missing electroweak corrections (EW). The negative sign indicates anti-correlation with respect to the unsigned uncertainties for SRs in the same column. Energy-related values are given in GeV.

SR	$n_j = 0$						$=1$
	Scale	PDF	Gen	EW	UE/PS	Tot	Tot
$e\mu/\mu e$, $10 < m_{\ell\ell} < 30$							
$p_T^{\text{sublead}} > 20$	0.7	0.6	3.1	−0.3	−1.9	3.8	7.1
$15 < p_T^{\text{sublead}} < 20$	1.2	0.8	0.9	0.7	1.7	2.6	3.9
$10 < p_T^{\text{sublead}} < 15$	0.7	1.0	0.4	1.2	2.2	2.8	5.4
$e\mu/\mu e$, $30 < m_{\ell\ell} < 55$							
$p_T^{\text{sublead}} > 20$	0.8	0.7	3.9	−0.4	−2.4	4.8	7.1
$15 < p_T^{\text{sublead}} < 20$	0.8	0.7	1.0	0.5	1.0	2.0	4.5
$10 < p_T^{\text{sublead}} < 15$	0.7	0.8	0.5	0.8	1.5	2.1	4.5
$ee/\mu\mu$, $12 < m_{\ell\ell} < 55$							
$p_T^{\text{sublead}} > 10$	0.8	1.1	2.4	0.1	−1.2	2.9	5.1

studied, while there is only one for the $ee/\mu\mu$ signal region.

The PDF uncertainties are evaluated using 68% C.L. CT10 [75] PDF eigenvectors and a comparison of the CT10 PDFs to the MSTW2008 and NNPDF2.3 PDFs. The envelope of all uncertainty variations is used. The QCD scale uncertainty is estimated following the standard procedure using aMC@NLO generator. The ME-UE/PS matching is checked by comparing POWHEG+Herwig and aMC@NLO+Herwig. The UE/PS uncertainty is extracted by comparing POWHEG+Pythia6, and POWHEG + Herwig. The above studies are done at truth level (also called generator level which is before detector simulation) with the validity of using truth samples checked. Except for the four common sources, the uncertainty of the NLO electroweak correction is estimated by reweighting the WW MC sample to the NLO EW calculation [76] and comparing with the original sample. All the results are summarised in Table 5.16 [49].

Another systematic error comes from the uncertainty of the cross section of the $gg \to WW$ process, which amounts to 5.8% (6.5%) of the total WW background in the signal region of $n_j = 0$ ($n_j = 1$) and 4.5% (3.7%) in the control region. This different fractions of $gg \to WW$ in the SR and CR may bias the extrapolation. An uncertainty is assigned to count for this difference, which is quantified by the difference of the fractions and is estimated to be 0.5% (3.1%) in the $n_j = 0$ ($n_j = 1$)

channel.

Since the WW is the main background in the $e\mu/\mu e$ $n_j = 0$ and $n_j = 1$ channels, the m_T shape uncertainty is also studied for the QCD scale, ME-UE/PS matching and UE/PS modelling, giving a maximum change of 20%.

WW in the $n_j \geq 2$ channels

In the $n_j \geq 2$ channels, MC predictions are used for both normalisation and shape. For the "QCD WW+2jets", the dominating uncertainty is from the QCD scale variation, which is 27% for the VBF BDT analysis and 19% for the ggF analysis. The second largest uncertainty is the ME matching and UE/PS uncertainty, which is evaluated by comparing Sherpa with Madgraph+Pythia6. The resulting uncertainty is output BDT (VBF) or m_T (ggF) bin dependent: 8–14% across BDT bins and 1–7% across m_T bins. For the "EW WW+2jets", the same procedures are carried out, resulting in a 10% normalisation uncertainty for both VBF and ggF analyses, 10–16% normalisation+shape uncertainty for VBF, and 5–17% normalisation+shape uncertainty for ggF. Besides, there could be an interference between the "QCD WW+2jets" and "EW WW+2jets" processes, which is estimated by Madgraph to be a few percent.

5.5.3 Top

Top in the $n_j = 0$ channel

The method of the top background estimation in the $n_j = 0$ channel is not the typical control region method (as shown in Formula 5.8) whose uncertainty can be assessed by studying the extrapolation factor. Rather, all top MC related systematics enter into the term $\frac{P_2^{\mathrm{MC}}}{(P_1^{\mathrm{MC}})^2}$ in Formula 5.19, which plays the role of the extrapolation factor. With this in mind, the theoretical systematic study can be performed as usual.

Truth level samples without detector simulation are used for the study, since it is impossible to have a reconstruction level sample with enough statistics being produced in a short time scale. Another argument is that we are studying theoretical uncertainty which should in principle not be affected by detector simulation. Anyway, this idea is checked by comparing distributions of the truth-sample with those of the reco-sample and the result is that they are roughly in agreement if excluding the detector effects.

The QCD scale uncertainty is studied using MC@NLO+Herwig and by varying the factorisation scale (μ_F) and renormalisation scale (μ_R). The variation of $\frac{P_2^{\rm MC}}{(P_1^{\rm MC})^2}$ in each case is shown in Table 5.17. The largest one is chosen as the uncertainty, which is: 0.7%.

Table 5.17: Relative variation (in %) of $\frac{P_2^{\rm MC}}{(P_1^{\rm MC})^2}$ in each case of varied factorisation scale (μ_F) and renormalisation scale (μ_R).

QCD scale	$\mu_F = 0.5$	$\mu_F = 1$	$\mu_F = 2$
$\mu_R = 0.5$	0.5 ± 0.2	0.5 ± 0.2	
$\mu_R = 1$	-0.4 ± 0.2	–	-0.7 ± 0.2
$\mu_R = 2$		-0.7 ± 0.2	-0.7 ± 0.2

The uncertainty of generator parton shower matching is estimated by comparing MC@NLO+Herwig with POWHEG+fHerwig to be 1.2%. For the UE/PS uncertainty, it is evaluated by comparing POWHEG+fHerwig with POWHEG+Pythia6 to be: 3.3%.

For the PDF uncertainty, as usual, CT10 eigenvectors, MSTW, and NNPDF are compared. The different samples with different PDF sets are not separately generated. Rather they are from the same nominal sample by the reweighting method. The CT10 eigenvectors are the PDF sets with one of the 26 parameters of the central CT10 PDF being varied up/down. They are summed up in the following way: first of all the total up ($\Delta X^+_{\rm max}$) and down ($\Delta X^-_{\rm max}$) variations are calculated by [77]:

$$\Delta X^+_{\rm max} = \sqrt{\sum_i^N [\max(X_i^+ - X_0, X_i^- - X_0, 0)]^2}, \quad (5.41)$$

$$\Delta X^-_{\rm max} = \sqrt{\sum_i^N [\max(X_0 - X_i^+, X_0 - X_i^-, 0)]^2}, \quad (5.42)$$

where X_i^+, X_i^- and X_0 are the up and down variation of parameter i and the nominal value, respectively. There are in total 52 variations, which are summarised in Fig. 5.20 in terms of relative variations with respect to the nominal one. The largest one of $\Delta X^+_{\rm max}$ and $\Delta X^-_{\rm max}$ is set to be the PDF uncertainty, which in this case takes the values of 1.6%

There could be interference between the $t\bar{t}$ and Wt processes. For example, some Feynman diagrams for the leading order Wt production are shown in Fig. 5.21 [78].

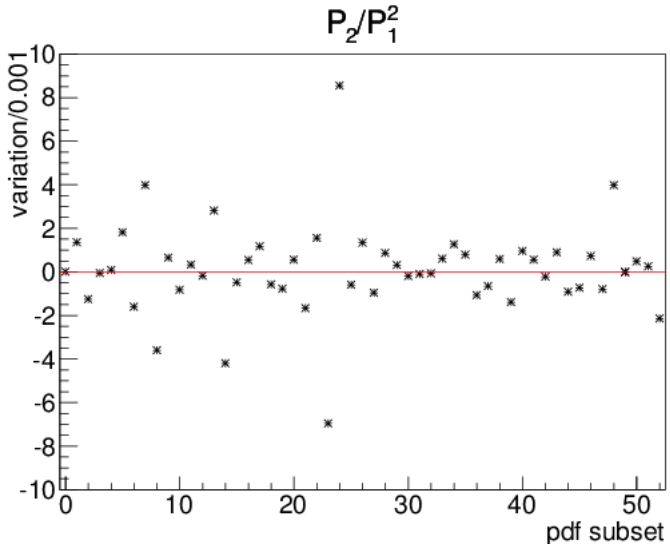


Figure 5.20: Relative variations in per mill on $\frac{P_2^{\mathrm{MC}}}{(P_1^{\mathrm{MC}})^2}$ as a function of CT10 PDF error eigenvectors based on a top MC sample at generator level. The first point represents the nominal PDF set and the other points correspond to 52 eigenvector variations.

They can also be interpreted as $t\bar{t}$ production with the one of the top being off-shell. To assess the interference uncertainty, two schemes of $t\bar{t}$ overlap removal for Wt production are implemented in MC@NLO: one scheme does the removal at the matrix element level called the DR scheme, and the other does it at the cross section level called the DS scheme [78]. By comparing these two schemes, the interference uncertainty can be estimated and the resulting value is 0.2%.

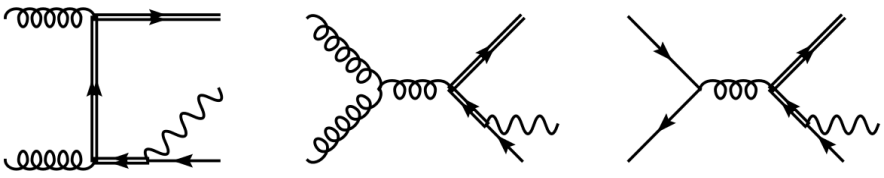


Figure 5.21: Wt production Feynman diagrams that have dual interpretations. Double lines represent the top quark.

There is one more systematic source, which is quite similar as the fraction uncertainty of $gg \to WW$ over all continuum WW backgrounds. That is the uncertainty on the relative contribution of the single top process in the top quark background. It is evaluated by varying the single top cross section up and down by 20%. The resulting uncertainty is 0.4%.

The above uncertainties are estimated at the jet veto level. The contribution of

the top quark background to the final signal region and WW $n_j = 0$ control region is computed by extrapolating the data-driven evaluated top at the jet veto to these regions. Thus the uncertainties of these extrapolation factors $\alpha^{MC}_{0j\to sr}$ and $\alpha^{MC}_{0j\to wwcr}$ should be estimated. All these uncertainties are summarised in Table 5.18.

Table 5.18: Summary of the theoretical uncertainties (in %) considered in the top background estimation based on the JVSP method at the zero jet-veto cut level.

Source	$\Delta\frac{P_2^{\rm MC}}{(P_1^{\rm MC})^2}$ %	$\Delta\alpha^{\rm MC}_{0j\to \rm SR}$ %	$\Delta\alpha^{\rm MC}_{0j\to WW\ \rm CR}$ %
PDF	1.6	0.8	2.7
Renormalisation and factorisation scales	0.7	1.1	0.8
Generator and UE/PS models	3.5	4.1	0.7
Relative Wt cross section variation ($\pm 20\%$)	$-0.4/0.3$	$-1.1/1.2$	$-0.4/0.4$
$t\bar{t}$ and Wt interference	0.2	0.2	0.2
Total	3.9	4.5	2.9

The m_T shape uncertainty is also investigated, which is done in two steps. First of all, it is found that the m_T shapes are similar between the top MC sample with at least one tagged b-jet and the sample without any tagged b-jet, which is shown in Fig. 5.22. Thus the MC predicted m_T shape uncertainty could be checked by comparing the data/MC agreement in the b-tagged control sample, which is shown in Fig. 5.23. The 4 plots from left to right and top to bottom correspond to inclusive, $n_j = 0$, $n_j = 1$ and $n_j \geq 2$ samples which are b-tagged. For the 0-jet sample, the b-jet p_T threshold has been lowered from the nominal value of 25 GeV to 20 GeV. Since we are interested only in the shape uncertainty, the normalisation of the total MC prediction has been fixed to that of data. As far as the 0-jet channel is concerned, given the limited event sample and a sizeable non-top contamination, it is hard to conclude whether there is any significant shape uncertainty on the top background.

Top in the $n_j = 1$ channel

For the top background estimation in the $n_j = 1$ channel, as shown in Formula 5.26, all the top MC related quantities are absorbed into the term $\alpha^{\rm MC}_{\rm eff}$, which is the extrapolation factor for the top tag efficiency from the $n_j = 2$ top control region. By calculating its variation, the uncertainties for the top background (after the bveto cut in the $n_j = 1$ signal region) can be estimated.

The systematic studies for $\alpha^{\rm MC}_{\rm eff}$ make use of reco-level samples here. The same set of theoretical sources are checked and the uncertainties are all found to be rather

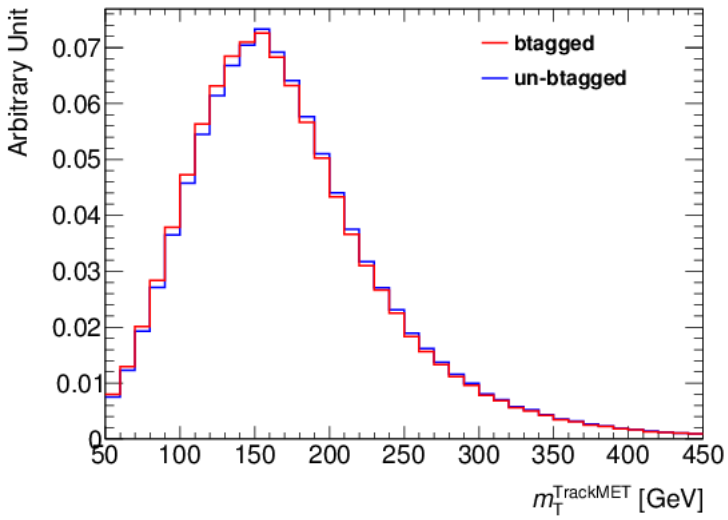


Figure 5.22: Comparison of the shape of the M_T distribution of the b-tagged sample (red histogram) with that of untagged sample (blue histogram).

small. Summing up all of them, we get a small theoretical uncertainty of 0.8% on $\alpha_{\text{eff}}^{\text{MC}}$ and if propagating to the final estimated top background, it is calculated to be 3.1%.

To get an estimation for the top events in the final signal region as well as in the WW control region (which is seriously contaminated by top events), the uncertainties of the extrapolation factors $\alpha_{\text{untag}\to\text{SR}}^{\text{MC}}$ and $\alpha_{\text{untag}\to WW\text{ CR}}^{\text{MC}}$ are studied in the common way, but this time with truth-level samples, which are the same sets of samples I produced for the JVSP systematics study. The results are summarised in Table 5.19 [79].

Table 5.19: Summary of the theoretical uncertainties (in %) of extrapolation factors from top after the b-jet veto to the final signal region and the WW control region in the $n_j = 1$ channel.

Source	$\Delta\alpha_{\text{SR}}$ %	$\Delta\alpha_{WW\text{ CR}}$ %
Scale	−1.1	0.6
PDF	−0.12	0.08
Matrix element	−2.4	2.0
Parton Shower	2.4	1.8

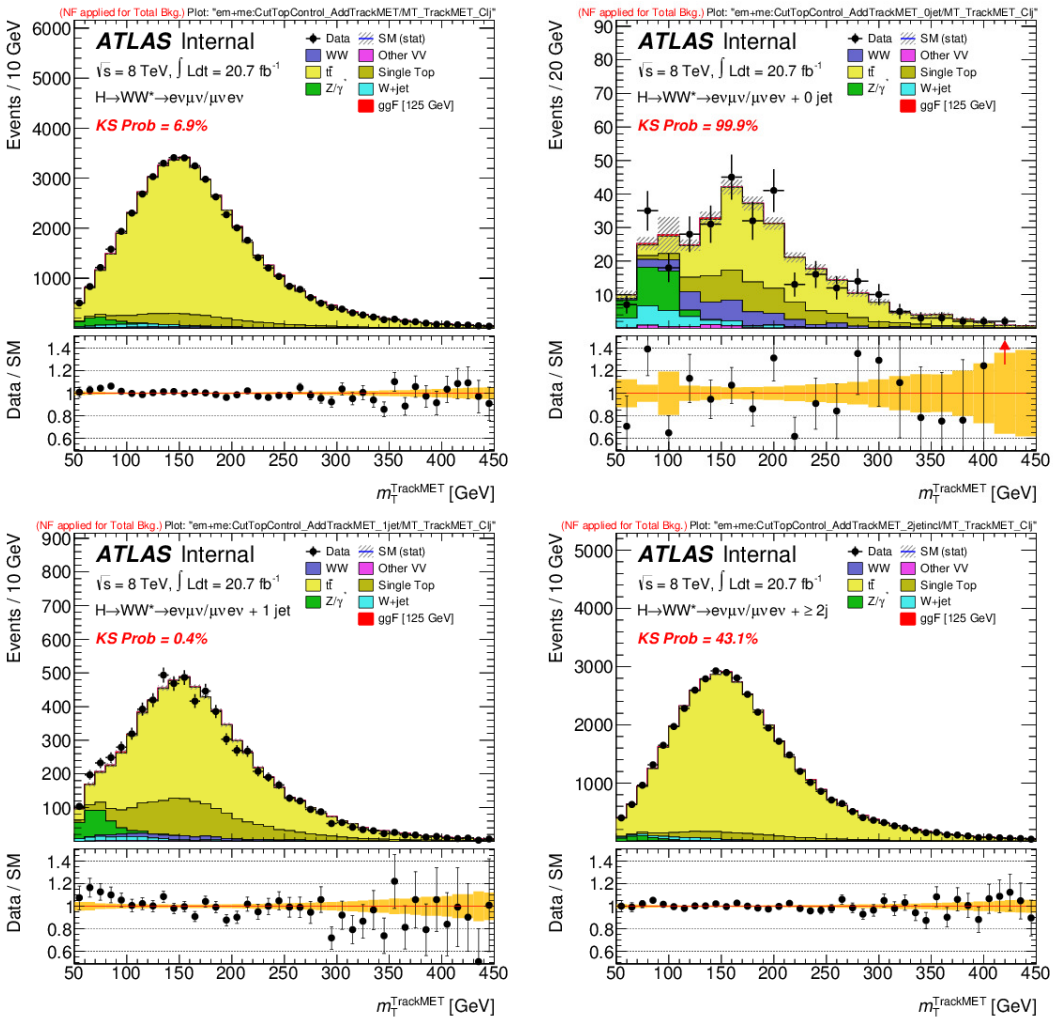


Figure 5.23: Distributions of M_T of the full b-tagged sample (top-left), 0-jet (top-right), 1-jet (bottom-left) and 2 or more jets (bottom-right) samples. The error band is statistical only.

Table 5.20: Top background uncertainties (in %) for VBF $n_j \geq 2$ channel on the extrapolation factors. The values together with the statistical and systematic uncertainties for the normalisation factors are also shown. The results of BDT bin 0 is unused, but also noted here for completeness.

BDT bins	$\Delta\alpha$ %	$\Delta\beta$(stat.)	$\Delta\beta$(sys.)	β
SR bin 0 (unused)	0.04	0.02	0.05	1.09
SR bin 1	0.10	0.15	0.55	1.58
SR bin 2	0.12	0.31	0.36	0.95
SR bin 3	0.21	0.31	0.36	

Top in the $n_j \geq 2$ channel

For the top background estimation in the $n_j \geq 2$ channels (ggF and VBF), the simple control region method is used, which means that theoretical uncertainties are estimated by studying the extrapolation factors. For the VBF analysis, there are two control regions (the BDT bin 1 and BDT bin 2+3) corresponding to two normalisation factors and three extrapolation factors (one for each high score BDT bin). After taking the standard procedures, theoretical uncertainties are calculated and summarised in Table 5.20 [49]. For the ggF analysis, everything is similar and the final total theoretical extrapolation uncertainty is 3.6%. Besides, the m_T shape modelling is checked for the ggF analysis and found to be at maximum 4% in the tail.

5.5.4 W+jets and QCD

These two backgrounds are estimated using the *Fake Factor method*, the theoretical uncertainty mainly goes into the correction factor for the fake factor which is MC-based rather into the fake factor which is data-driven. This correction factor actually can also be treated as an extrapolation factor which extrapolates the fake factor from the Z+jets sample to the W+jets sample.

The theoretical uncertainties of the correction factors are extracted by comparing Alpgen+Pythia6, Alpgen+Herwig and POWHEG+Pythia8 generator samples (so it is in fact the generator and UE/PS uncertainty). The results are: 20% (22%) for the fake electron (muon) in the opposite sign (OS) region (signal region) and 25% (35%) in the same sign (SS) region (control region). Besides, the systematics are partially correlated between the OS and SS regions due to the asymmetrical possibility of jets being misidentified as SS and OS leptons. The QCD theoretical

uncertainties are estimated in a similar way.

Here we only considered the uncertainty of the correction factor, which should also be done for the fake factor, whose uncertainties are mainly from the low statistics of the jet samples and the subtraction of the EW contamination using MC. However, since they are treated as experimental uncertainty in the fit so will be discussed in later sections.

5.5.5 Non-WW dibosons

For the normalisation of non-WW diboson backgrounds in the $e\mu/\mu e$ $n_j \leq 1$ channels, control regions are used to constrain the total non-WW dibosons, so in principle the uncertainties reside in the extrapolation factors. However, it is varied in the simulation that the related uncertainty is negligible.

Although the total rate is constrained, the individual diboson processes are able to float relatively. And they rely on MC predictions, thus suffer from theoretical systematics.

For the $W\gamma$ process, a k-factor of 1.15 is applied from the MCFM calculation with an overall 6% uncertainty and an exclusive jet bin uncertainty to be 9%, 53%, and 100% for $n_j = 0$, $n_j = 1$, and $n_j \geq 2$. These are from repeating the QCD scale variation, which is the dominant source. Besides, there is a 3% PDF uncertainty.

$W\gamma^*$ is treated similarly. The k-factor is calculated by MCFM to be 0.94 with a 7% overall uncertainty and 7%, 30%, and 26% jet bin uncertainty for $n_j = 0$, $n_j = 1$, and $n_j \geq 2$ channels. For WZ, since it is produced by POWHEG which is in NLO precision, no additional corrections or systematics are applied. Besides, the acceptance uncertainty is studied and found to be negligible.

5.5.6 Drell-Yan

For the $Z/\gamma^* \to \tau\tau$ which contributes mainly to the $e\mu/\mu e$ channel, the simple control region method is used. So the theoretical uncertainties are evaluated on the extrapolation factor from the CR to the SR following the standard procedures. Three usual uncertainties are studied: the QCD scale, PDF and the PS/UE. There is one more uncertainty which only exists for $n_j = 0$ channel: the Z/γ^* p_T reweighting uncertainty, which is to correct the mis-modelling of soft jets. This p_T reweighting is applied to all four lepton flavour channels. Besides, the extrapolation uncertainties from $Z/\gamma^* \to \tau\tau$ control regions to WW control regions are also studied as usual. All the results are summarised in Table 5.21 [49].

Table 5.21: $Z/\gamma^* \to \tau\tau$ uncertainties (in %) on the extrapolation factor (Scale), PDF and generator modelling (Gen) uncertainties are reported. For the $n_j = 0$ category, additional uncertainty due to the p_T^{Z/γ^*} reweighting is shown. The negative sign indicates anti-correlation with respect to the unsigned uncertainties in the same column.

Regions	Scale	PDF	Gen	p_T^{Z/γ^*}
Signal regions				
$n_j = 0$	−1.6	1.4	5.7	19
$n_j = 1$	4.7	1.8	−2.0	-
$n_j \geq 2$ ggF	−10.3	1.1	10.4	-
WW control regions				
$n_j = 0$	−5.5	1.0	−8.0	16
$n_j = 1$	−7.2	2.1	3.2	-

As for the $Z/\gamma^* \to \tau\tau$ estimation in the VBF channel, no systematic uncertainty is assigned since the statistical uncertainty is large and dominating, which can be seen from the calculated normalisation factor: 0.9 ± 0.3(stat.).

For the Drell-Yan estimation in the $n_j \leq 1$ same flavour channels, the Pacman method is used. Theoretical systematics lie in the various templates, which in real application are f_{recoil} ($f_{\text{recoil}}^{\text{extended}}$) cut efficiencies. These efficiencies are all extrapolated from control regions, so the problem simplifies into the uncertainties of extrapolation, which is evaluated by comparing cut efficiency differences between the $e\mu/\mu e$ and $ee/\mu\mu$ channels using MC simulation in different cases. The results are summarised in Table 5.22 [49]. The largest uncertainties are on ϵ_{DY} but since the non-DY component dominates in the signal region, the uncertainties on its cut efficiencies are the dominant contribution to the final total uncertainty on the estimated Drell-Yan.

The $Z/\gamma^* \to ee/\mu\mu$ in the VBF BDT analysis is estimated using the ABCD method. The correction factor $f_{\text{correction}}$ is treated as a non-closure factor, of which the difference with unity of 17% is taken as the non-closure uncertainty. The method assumes no correlation between E_T^{miss} and the output BDT, to cover the possible bias, a systematic is assigned for the assumption across each to-be-fitted BDT bins: 4%, 10%, and 60% with increasing BDT score. Besides, there is a 11% uncertainty for the MC predicted shape in the merged two highest BDT bins (due to lack of statistics for the highest score bin) which is derived following the standard procedures of the theoretical systematic estimation.

Table 5.22: The systematic uncertainties for $Z/\gamma^* \to ee/\mu\mu$ background estimation in the $n_j \leq 1$ channels as well as the uncertainties on the various efficiencies.

Source	$n_j = 0$	$n_j = 1$
Uncertainty on $\epsilon_{\text{non-DY}}$	1.9	3.2
From statistical	1.8	3.0
From using $e\mu/\mu e$ CR to extrapolate to the SR	0.8	1.2
Uncertainty on ϵ_{DY}	38	32
From statistical	9.4	16
From using Z peak to extrapolate to the SR	32	16
Uncertainty on $\epsilon_{\text{non-DY}}^{Z \text{ peak}}$	3.1	4.5
From statistical	1.9	3.9
From using $e\mu/\mu e$ CR to extrapolate to the SR	2.5	2.4
Total uncertainty on Drell-Yan yield in SR	49	45

5.5.7 Others

For the DPI process, its cross section is not extremely well known, for the time being, a 60% uncertainty is assigned.

For the 7 TeV analysis, the theoretical uncertainties on the extrapolation factors used in the WW, top, and $Z/\gamma \to \tau\tau$ background estimation methods are assumed to be the same as in the 8 TeV analysis. Others are estimated using the same procedures as in the 8 TeV analysis.

5.6 Experimental systematics

Experimental systematics comes from imperfect experimental instruments. The size of the error is determined by the uncertainty of the calibration and resolution of the apparatus. For high energy physics, the experimental uncertainty lies in the measurement of the physical quantities like the deposit energy and time of flight, the understanding of the experiment environment, and the algorithms that are used to reconstruct physical objects out of electronic signals.

In the following, experimental uncertainties considered in the analysis will be introduced.

5.6.1 Lepton

The leptons (electron and muon) used in the analysis are already described in previous sections. In short, they are reconstructed from scratches and selected by

a series of cuts. Experimental uncertainty arises in each of the steps.

For electrons, it is reconstructed and identified first by the corresponding softwares, whose uncertainties are presented in Table 5.23 [37]. Then the resolution and energy scaling is measured together with their uncertainties, which are also shown in Table 5.23. These numbers are all from e/gamma group measurements. These electrons are further selected to meet the analysis needs by the impact parameter cuts, isolation requirements, etc. Each selection will potentially bring in additional systematics, which are evaluated by the Z tag-and-probe method. The estimated systematics are summaiised in Table 5.24 [37], where the total efficiency of all the selection and its uncertainties are E_T dependent.

Table 5.23: Electron systematic input from the e/gamma group.

Source of Uncertainty	Size of the uncertainty
Electron Efficiency	reconstruction : 0.1–1.0 % depending on E_T and η
	identification : 0.2–2.7 % depending on E_T and η
Electron Energy Scale	$\sim$ 0.4 % depending on E_T and η (except for crack region)
Electron Energy Resolution	about 1 % depending on E_T and η

Table 5.24: Total electron selection efficiencies and uncertainties for an m_H = 125 GeV Higgs signal sample. All uncertainties are added in quadrature for the total. (Energy scale and resolution are not included).

E_T (GeV)	Total Eff.	Iso. Unc. (relative)	ID+Rec. Unc. (relative)	Total Unc. (relative)
10-15	0.412	0.016	0.016	0.022
15-20	0.619	0.009	0.024	0.025
20-25	0.668	0.008	0.027	0.028
25-30	0.755	0.007	0.014	0.016
30-35	0.770	0.007	0.005	0.009
35-40	0.796	0.006	0.003	0.007
40-45	0.798	0.006	0.002	0.006
45-50	0.813	0.006	0.002	0.006

An additional systematic uncertainty is applied to the electrons reconstructed from converted photons, which is derived in a $Z + \gamma$ validation region and takes the value of $\pm 25\%$ for $10 < E_T < 15$ GeV, $\pm 0.18\%$ for $15 < E_T < 20$ GeV, and $\pm 5\%$ for $E_T > 20$ GeV.

For muons, there are similar systematic sources and they are summarised in Table 5.25 [37] for the resolution, scale and reco+id efficiency, which are from the MCP group. Systematics for the additional cuts are summarised in Table 5.26 [37].

Table 5.25: Muon systematic input from the MCP group.

Source of Uncertainty	Size of the uncertainty
Muon Efficiency	< 0.46 % depending on p_T and η
Muon Energy Scale	< 0.50 % depending on p_T and η
Muon Energy Resolution	less than 1 % depending on p_T and η

Table 5.26: Total muon selection efficiencies and uncertainties for an $m_H = 125\,\text{GeV}$ Higgs signal sample. All uncertainties are added in quadrature for the total. (Momentum scale and resolution are not included).

E_T (GeV)	Total Eff.	Iso. Unc. (relative)	ID+Rec. Unc. (relative)	Total Unc. (relative)
10-15	0.574	0.027	< 0.005	0.027
15-20	0.808	0.012	< 0.005	0.013
20-25	0.904	0.007	< 0.005	0.009
25-30	0.924	0.006	< 0.005	0.008
30-35	0.932	0.006	< 0.005	0.008
35-40	0.942	0.005	< 0.005	0.007
40-45	0.943	0.005	< 0.005	0.007
45-50	0.944	0.005	< 0.005	0.007

5.6.2 Jets

The experimental systematics for jets concerns mainly their energy and resolution: jet energy scale (JES) and jet energy resolution (JER), which is provided by the JetMET combined performance group covering all systematics emerging from the jet calibration and resolution.

12 parameters are used to parametrise the systematics on JES, which are:

1. in-situ: Detector1,
2. in-situ: Modelling1,
3. η intercalibration: modelling,
4. η intercalibration: stat+method,
5. high p_T jets,
6. in-time pile-up,
7. out-of-time pile-up,
8. pile-up p_T,

9. pile-up ρ topology,

10. flavour composition,

11. flavour response,

12. b-JES.

The JES uncertainty is p_T and η dependent and ranges from 1% to 7%. For the central region and forward region, the uncertainty is shown over p_T in Fig. 5.24 [37].

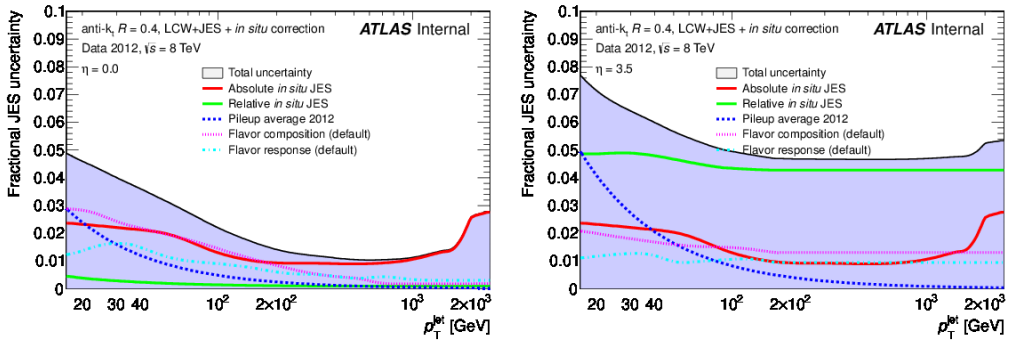


Figure 5.24: JES uncertainty in the central ($|\eta| = 0.0$, left) in the forward ($|\eta| = 3.5$, right) region for the LCW+JES jet collections.

To have a sense of which components are important, an inclusive analysis is performed to assess the impact of JES uncertainties on the signal (ggF and VBF) and $t\bar{t}$ samples, with the three dominant sources being shown in Table 5.27 [37].

Table 5.27: Leading JES systematics uncertainties at the preselection cut stage of the $e\mu/\mu e$ analysis.

Process	$n_j = 0$	$n_j = 1$	$n_j \geq 2$
ggF signal	FlavComp 2.2%	η model 1.0%	η model 12.6%
VBF signal	η model 9.0%	η model 3.9%	η model 4.9%
$t\bar{t}$	InSitu Model 8.2%	InSitu Model1 4.6%	η model 6.0%

Systematics on JER is one independent source also provided by the JetMET CP group. It is also p_T and η dependent and varies between 2% to 40%, with the worst uncertainty occurring around the p_T threshold of the jet selection.

There are other two jet related systematics to be considered:

- Pile-up induced migration uncertainty. The pile-up induced migrations are defined as the probability that the event belonging to jet bin A migrates to jet bin B (B = A+1) due to a well reconstructed pile-up jet,

- JVF uncertainty from data/MC discrepancy on the JVF cut efficiency, mainly through the CJV cut in the VBF channel (∼0.5%).

All the systematics are summarised in Table 5.28 as well as their treatments in the analysis.

Source	$n_j = 0$	$n_j = 0$ treatment	$n_j = 1$	$n_j = 1$ treatment	VBF	VBF treatment
$n_j = 0 \to n_j = 1$	$\sim 0.5\%$ (signal)	neglected	x	N/A	x	N/A
$n_j = 1 \to n_j \geq 2$	x	N/A	$\sim 0.5\%$ (signal)	neglected	x	N/A
$n_j = 1 \to$ VBF	x	N/A	x	N/A	$\sim 0.05\%$ (signal)	neglected
$n_j \geq 2 \to$ VBF CJV	x	N/A	x	N/A	1% (signal)	added to the sys fwk by hand
VBF CJV JVF	x	N/A	x	N/A	$< 0.5\%$ (signal)	neglected
leading JES (signal)	2.2%	automatic by JES Unc. Provider	1.0%	automatic by JES Unc. Provider	4.9%	automatic by JES Unc. Provider
leading JES (top bkg)	8.2%	automatic by JES Unc. Provider	4.6%	automatic by JES Unc. Provider	6.0%	automatic by JES Unc. Provider

Table 5.28: Jet related systematics summary and the treatment in the analysis.

5.6.3 *b*-tagging

The uncertainties related to *b*-tagging include [37]:

- Uncertainty of the *b*-jet identification, which is decomposed into 6 uncorrelated components using the eigenvector method which ranges between <1% to 7.8%,
- Uncertainty of misidentified light jets as *b*-jets, which is p_T and η dependent and ranges between 9% and 15% in the central region ($|\eta| < 1.1$) and between 9% and 19% in the forward region ($1.1 < |\eta| < 2.5$),
- Uncertainty of misidentified *c*-jets as *b*-jets, which is p_T dependent and ranges between 6% and 14%.

5.6.4 Pile-up

The pile-up systematics has three impacts on the analysis [37]:

1. Some of its effects are parametrised into the JES parameters: pile-up p_T and ρ with maximum impact of $\sim$ 2%,
2. It influences the μ rescaling up/down by 0.1: from 0.9 (nominal) to 0.8 and 1.0 with worst impact to be 2–4%,

3. Appearance of hard pile-up jet migrates events between jet bins, which mainly affects the VBF analysis with a 1% uncertainty.

5.6.5 Missing transverse energy

The systematics estimation of the E_T^{miss} concerns only the soft term, since the systematics of the hard terms, which are involved in the calculation of the E_T^{miss}, are propagated into the systematics of the E_T^{miss} automatically.

For the calo-based $E_T^{\text{miss, calo}}$, the soft term $E_T^{\text{miss, CellOut}}$ is decomposed to longitudinal (L) and perpendicular (P) directions with respect to the hard components. The systematics are obtained by smearing and scaling both components:

- Shifting up or down the $< E_{T,L}^{\text{miss, CellOut}} >$, which is the scale variation,
- Scaling $E_{T,L}^{\text{miss, CellOut}}$ and $E_{T,P}^{\text{miss, CellOut}}$ simultaneously up or down assuming the two quantities are correlated,
- Scaling $E_{T,L}^{\text{miss, CellOut}}$ and $E_{T,P}^{\text{miss, CellOut}}$ in opposite direction simultaneously assuming the two quantities are anti-correlated.

For the track-based jet corrected $E_T^{\text{miss, track, jetCorr}}$, a similar decomposition of the soft term $E_{T,\text{soft}}^{\text{miss, track, jetCorr}}$ is done resulting into six variations:

- Scaling up or down $E_{T,\text{soft}}^{L,\text{miss, track, jetCorr}}$,
- Smearing and scaling up or down $E_{T,\text{soft}}^{L,\text{miss, track, jetCorr}}$,
- Smearing up or down $E_{T,\text{soft}}^{T,\text{miss, track, jetCorr}}$.

The track-based non-jet corrected $E_T^{\text{miss, track}}$ shares the same systematics variation with $E_T^{\text{miss, track, jetCorr}}$ as above in terms of soft terms and has additional uncertainties from the tracks inside the jets.

For the calo-based $E_T^{\text{miss, calo}}$ soft terms, the resulting scale uncertainty of the longitudinal component ranges between 0.2 to 0.3 GeV. And the resolution of both the longitudinal and perpendicular components varies between 1% to 4%. All the uncertainties of the soft terms are hard-object p_T and pile-up dependent. While for the track-based $E_T^{\text{miss, track}}$ soft terms, the resulting scale uncertainty of the longitudinal component ranges between 0.3 to 1.4 GeV. And the resolution of both the longitudinal and perpendicular components varies between 1.5 GeV to 3.3 GeV. All the uncertainties of soft terms are hard-object p_T dependent.

5.6.6 Luminosity

The uncertainty on the integrated luminosity in the 8 (7) TeV analysis is 2.8% (1.8%) [80]. It is derived from a preliminary calibration of the luminosity scale derived from beam-separation scan [81].

5.6.7 Summary

Unlike the theoretical uncertainty which is region dependent and process dependent, the experimental uncertainty mentioned above influence all the parts of the analysis in a coherent way. It has impact on both the normalisations and shapes. But it is found that the shape variation induced by the experimental systematics is not obvious in most cases and not important even if it changes the variable shape significantly.

5.7 Statistics framework

With the event selection described previously, the initial data sample is now divided into several subsamples in which the signal sensitivities are maximised thanks to the optimised cuts. In the phase spaces of each subsamples, the contributions of signals and backgrounds have been calculated by either Monte Carlo or data-driven methods.

All we have to do now is to test quantitatively if the predicted signal is compatible with the observation or equivalently to measure the signal strength μ, which is defined as the normalised signal event rate over the SM prediction. Thus a theory with $\mu = 1$ is the SM, and with $\mu = 0$ is the non-Higgs SM. For the aim of this chapter, which is the observation of the SM Higgs, two goals can be set:

- Rejecting the non-Higgs model,
- Measuring the signal strength and its compatibility with the SM.

The basic mathematic tools used to reach these goals are introduced in this section, which establish the likelihood-based statistical analysis. Then the likelihood construction is specified.

5.7.1 Likelihood and statistical test

A likelihood $\mathcal{L}$ is a function of the parameters of a statistical model: $\mathcal{L}(\mu|x)$, where μ represents all the parameters and x the outcome of the experiment. It has the same form as the probability density function (PDF) of the expected experiment outcome $f(x|\mu)$, which depends on the parameters of the model. But it should be distinguished from the PDF: PDF is in a sense an absolute measure of the probability of the outcome x, while the Likelihood only makes sense when compared between several points of the parameter space and it gives a tendency to which point the outcome data supports.

To find the best value of the parameter μ of the model that is most compatible with the data, a likelihood scan can be performed over the parameter space and the best value, denoted as $\hat{\mu}$, corresponds to the point where the likelihood reaches a maximum. Besides, the likelihood method makes the combination of different measurements easy. In our case, the 7 TeV and 8 TeV measurements can be simply merged by multiplying their likelihoods together.

In practice, not all parameters of the model are of interest to us, so they are separated out in the likelihood $\mathcal{L}(\mu, \theta|x)$, where μ represents now those interesting parameters and is called "parameters of interest" (POIs) and θ those other parameters and called "nuisance parameters" (NPs). These nuisance parameters can be eliminated by many approaches. In this analysis, the profile likelihood method is used, which reduces NPs by writing them as functions of the POIs $\hat{\hat{\theta}}(\mu)$, whose function form is derived by maximising the likelihood function giving the POIs fixed.

Besides, the profiled likelihood is always normalised to its maximum value $\mathcal{L}(\hat{\mu}, \hat{\theta}))$ given by the unconditional maximum-likelihood (ML) estimator $\hat{\mu}$ and $\hat{\theta}$, which is called profile likelihood ratio and takes the form $\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\theta})}$. In a real application, since the signal strength is defined to be non-negative, the profile likelihood ratio is constructed as:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\theta})} & \hat{\mu} \geq 0\,, \\ \frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0\,, \end{cases} \tag{5.43}$$

based on which, the test statistic can be defined as:

$$\tilde{t}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0\ , \\ -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\theta})} & 0 \leq \hat{\mu} \leq \mu\ , \\ 0 & \hat{\mu} > \mu\ . \end{cases} \tag{5.44}$$

It is used to measure the disagreement between the data and the model characterised by the model parameter μ. The larger the observed $\tilde{t}_{\mu,\text{obs}}$ is, the larger the disagreement exists between the data and the model. The test $\tilde{t}_\mu$ is a function of the observed data as it is derived from the likelihood. Thus it is a random number with a certain PDF. Usually, we feed pseudo random data which is sampled from a known hypothesis μ to the test, then we have the sample distribution $f(\tilde{t}_\mu|\mu)$. Knowing the sample distribution, we can accurately quantify the extent of the disagreement by the integral of $\tilde{t}_\mu$ from the point of observed $\tilde{t}_{\mu,\text{obs}}$ to infinity:

$$p_\mu = \int_{\tilde{t}_{\mu,\text{obs}}}^{\infty} f(\tilde{t}_\mu|\mu)d\tilde{t}_\mu\ . \tag{5.45}$$

This is the so-called "p" value for the hypothesis μ, the smaller it is, the larger the disagreement is. The calculation is also illustrated in Fig. 5.25(a) [82], where the function shown is the sample distribution of test $\tilde{t}_\mu$ under the assumption of hypothesis μ, and the area in blue (shaded) is the p-value.

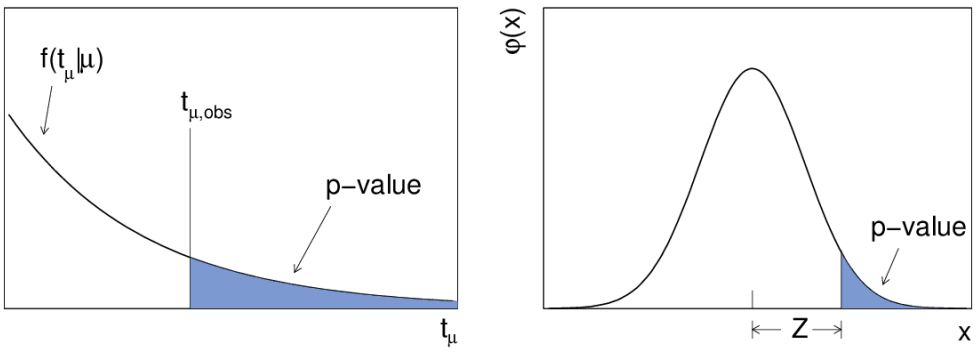


Figure 5.25: Illustration of the calculation of "p" value and the corresponding statistical significance "Z".

In case of rejecting the non-signal theory, μ takes the value of 0 both for the test and the assumed theory to drive the sample distribution. So the "p_0" value of

background-only hypothesis is defined as:

$$p_0 = \int_{\tilde{t}_{0,\text{obs}}}^{\infty} f(\tilde{t}_0|0)d\tilde{t}_0 \,. \tag{5.46}$$

There is a probability interpretation of the "p" value: it quantifies the probability that the observed signal is from a statistical fluctuation of the backgrounds described by the background-only theory. A more convenient quantity called statistical significance can be defined based on the p-value by translating from the Gaussian tail probability:

$$Z = \Phi^{-1}(1 - p_0) \,, \tag{5.47}$$

where Φ^{-1} is the quantile of the standard Gaussian and Z quantifies the number of standard deviation. The calculation is also schematically shown in Fig. 5.25(b) [82].

Besides, there will be the requirement to estimate the upper limit of the parameter of interest. In this case, the 95% confidence level upper limit on μ can be inferred using the so-called CLs method. The CLs is defined by:

$$\text{CLs} = \frac{p_\mu}{1 - p_b} \tag{5.48}$$

where p_μ is the p-value defined in Formula 5.45, and p_b is similar as p_μ except that the sampling distribution is derived from the null hypothesis of $\mu = 0$. The value of μ that makes the CLs equaling to 0.05 is the upper limit.

5.7.2 Likelihood construction

From the above section, we know that to draw a statistical conclusion, the most important quantity to start from is the likelihood function. For the $H \to WW^* \to \ell\nu\ell\nu$ analysis, the fit goes through many regions with many systematic constraints. In the following, the likelihood will be constructed step by step.

The fit is carried out in each bin of the variable distribution (m_T or output BDT) in each signal region. Among all signal regions, each of the two in the $e\mu/\mu e$ $n_j = 0$ and $n_j = 1$ channels is further divided into 12 sub-regions according to $m_{\ell\ell}$, p_T^{sublead}, and sub-leading lepton flavour (e or μ). So there are in total 29 signal regions. In the bin b of the region i, the probability of observing N_{ib} events given

λ_{ib} expected events is modelled by a Poisson distribution:

$$P(N_{ib}|\lambda_{ib}) = e^{-\lambda_{ib}}\lambda_{ib}^{N_{ib}}/N_{ib}!\,, \tag{5.49}$$

where λ_{ib} is the sum of expected signal and backgrounds:

$$\lambda_{ib} = \mu \cdot S_{ib} \cdot \prod_{p}^{P} \nu_{bp}(\theta_p) + \sum_{k} \beta_k \cdot B_{kib} \cdot \prod_{q}^{Q} \nu_{bq}(\theta_q)\,. \tag{5.50}$$

In the above formula, μ is the defined signal strength, the parameter of interest (However, it should be mentioned that there could be more than one POI). S_{ib} and N_{kib} are the expected signal and type k background in signal region i bin b, while β_k is the normalisation factor for the k background. The variable ν is a response function that parametrises the impact of the systematic uncertainty θ on the signal or background yields, with p and q being indices for the signal and background systematics sources. Correspondingly, P and Q are the collections of systematics. These systematics are constrained by a nominal Gaussian function:

$$G(0|\theta, 1) = e^{-\theta^2/2}/\sqrt{2\pi}\,. \tag{5.51}$$

Their response functions ν_{bp} and ν_{bq} take the form of $(1+\epsilon)^\theta$ when the uncertainty ϵ is universal across each bin b and $(1+\epsilon_b \cdot \theta)$ when the uncertainty affects the distribution over each b by ϵ_b. To constrain the nuisance parameter β_k, either an input from outside is given or a control region for background k enters the fit. In the former case, e.g. the top background in the $n_j = 0$ channel, the normalisation factor β and its uncertainty ϵ are calculated prior the fit. In the latter case, a Poisson term that describes the observed event count N_l in the control region l (for controlling background l) is added into the likelihood:

$$P(N_l|\lambda_l)\,, \tag{5.52}$$

where λ_l is the expected event rate in the control region, which is represented as:

$$\beta_l \cdot B_{ll} + \sum_{k\neq l} \beta_k \cdot B_{kl} \cdot \prod_{s} \nu_s(\theta_s))\,. \tag{5.53}$$

In the equation, B_{kl} is the expected event yield of background k in the control region for the background l, $\nu_s(\theta_s)$ is the realisation of the same systematics s

appeared in the signal region.

Besides, there is another type of systematics which is caused by the finite size of the MC sample N_{MC} used to predict the expected event yields. They are constrained by a Poisson function instead of Gaussian:

$$P(N_{\mathrm{MC}}|N_{\mathrm{MC}}\cdot\theta)\,. \tag{5.54}$$

Such a systematics exists in every bin/region used in the fit wherever the MC samples entered. And the response function $\nu(\theta)$ takes a simple form of $\nu(\theta)=\theta$.

Now we get the final likelihood by unifying all the above formulas and multiplying together all the fit regions and constraints:

$$\mathcal{L}=\prod_{ib}^{sr,bin}P(N_{ib}|\lambda_{ib})\cdot\prod_{l}^{cr}P(N_l|\lambda_l)\cdot\prod_{t}^{P,Q}G(0|\theta_t,1)\cdot\prod_{n}^{sr,bin,cr}P(N_{MC,n}|N_{MC,n}\cdot\theta_n) \tag{5.55}$$

5.8 Results

By applying the statistical tools introduced in the previous section, like the likelihood fit, profile likelihood fit, statistic test, and CLs, to the likelihood function 5.55, prolific conclusions can be drawn from the observed data.

5.8.1 Observation of the Higgs boson

By using the p_0 defined in Formula 5.46 to test the background-only SM ($\mu=0$), the observed excess of data over the expected backgrounds, which is the signal significance, can be quantified. The observed and expected p-values are shown as a function of the Higgs mass m_H in Fig. 5.26 [49]. From the plot, it is observed that the local minimum of p_0 lies at $m_H\simeq 130\,\mathrm{GeV}$, corresponding to a statistical significance of $6.1\,\sigma$. The expected significance at that point is $5.8\,\sigma$. This result rejects the background-only theory and establishes a discovery-level signal ($>5\,\sigma$) in the $H\to WW^*\to\ell\nu\ell\nu$ channel alone.

The observed signal can be interpreted as the SM Higgs. Since the Higgs mass is a free parameter in the SM, the compatibility of each Higgs mass (which corresponds to a different model) with the signal can be assessed by looking at the best-fit signal strength $\hat{\mu}$. The result is shown in Fig. 5.27 [49]. The observed $\hat{\mu}$ is close to unity for the Higgs mass of $\sim 125\,\mathrm{GeV}$, which implies this mass point is most compatible with the observation. More accurately, if setting $m_H=125.36\,\mathrm{GeV}$,

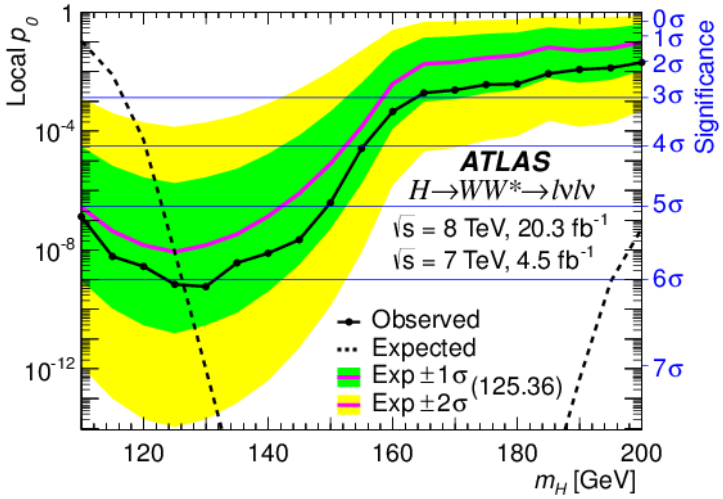


Figure 5.26: Local p_0 value as a function of m_H. The observed values are shown as a solid line with points where p_0 is evaluated. The dashed line shows the expected values given the presence of a signal at each x-axis value. The expected values for $m_H = 125.36\,\text{GeV}$ are given as a solid line without points; the inner (outer) band shaded darker (lighter) represents the one (two) standard deviation uncertainty.

the observed best-fit of μ takes the value:

$$\begin{aligned} \mu &= 1.09^{+0.16}_{-0.15}\,(\text{stat.})^{+0.08}_{-0.07}\,(\text{exp.})^{+0.15}_{-0.12}\,(\text{theo.}) \pm 0.03\,(\text{lumi.}) \\ &= 1.09^{+0.16}_{-0.15}\,(\text{stat.})^{+0.17}_{-0.14}\,(\text{syst.}) \\ &= 1.09^{+0.23}_{-0.21}\,, \end{aligned} \tag{5.56}$$

while the expected one is: $1\,^{+0.16}_{-0.15}\,(\text{stat.})\,^{+0.17}_{-0.13}\,(\text{syst.})$.

Besides, if relaxing the relationship between m_H and the total event yield, then we can let both m_H and μ to float so that the two-dimension fit of (m_H, μ) can be performed. The result is shown in Fig. 5.28 [49]. The value ($\mu = 1$, $m_H = 125.36\,\text{GeV}$), which represents the Higgs mass measurements in the high-resolution channels [83], lies well within the 68% C.L. contour, showing that the observation here is compatible with those measurements.

Rather than checking the compatibility of each Higgs mass point to the excess, the interpretation could go in the opposite way: using the CLs method introduced before to exclude @95% C.L. those incompatible points. The same mass range as shown in Fig. 5.27 is examined, with exclusion limits shown in Fig. 5.29 [49].

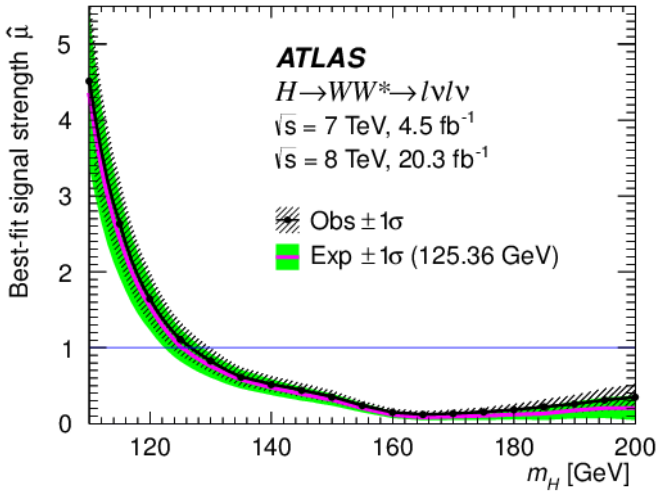


Figure 5.27: Best-fit signal strength $\hat{\mu}$ as a function of m_H. The observed values are shown as a solid line with points where $\hat{\mu}$ is evaluated. The expected values are shown as a solid line without points. The dashed and shaded (solid) bands represent the one standard deviation uncertainties for the observed (expected) values.

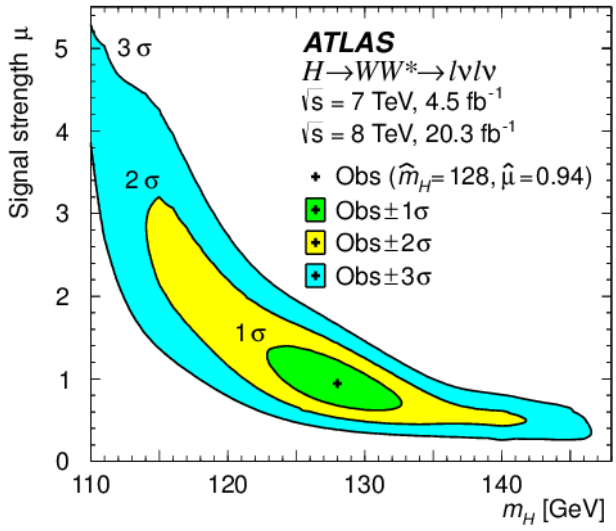


Figure 5.28: Observed signal strength μ as a function of m_H as evaluated by the likelihood fit. The shaded areas represent the one, two, and three standard deviation contours with respect to the best fit values $\hat{m}_H$ and $\hat{\mu}$.

5.8.2 Evidence for VBF

As mentioned before, we can split the signal strength μ into the ggF signal strength μ_{ggF} and the VBF signal strength μ_{VBF} and do a ggF-profiled fit to extract the significance of the VBF signal. In practice, the parameter used to do the

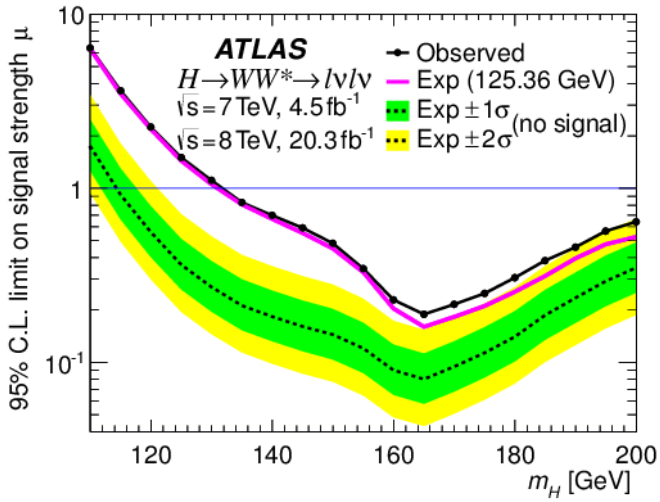


Figure 5.29: CLs exclusion plot for $110 < m_H < 200\,\text{GeV}$. The observed values are shown as a solid line with points where the limit is evaluated. The expected values for a signal at 125.36 GeV are given as a solid line without points. The expected values for scenarios without signal are given by the dotted line. The inner (outer) band shaded darker (lighter) represents the one (two) standard deviation uncertainty on the value for expected without signal. The limit of 132 GeV (114 GeV) on μ for the observed (expected no signal) scenario can be seen at low values of m_H.

likelihood scan is the VBF signal strength normalized to that of ggF $\frac{\mu_{\text{VBF}}}{\mu_{\text{ggF}}}$, which is shown in Fig. 5.30 [49]. From the plot, the best-fit of the signal strength ratio is [49]:

$$\frac{\mu_{\text{VBF}}}{\mu_{\text{ggF}}} = 1.26\,^{+0.61}_{-0.45}\,(\text{stat.})\,^{+0.50}_{-0.26}\,(\text{syst.}) = 1.26\,^{+0.79}_{-0.53}\,. \tag{5.57}$$

The value of the likelihood at the point $\frac{\mu_{\text{VBF}}}{\mu_{\text{ggF}}} = 0$ can be interpreted as the statistical significance of the VBF signal over non-VBF hypothesis, corresponding to an observed (expected) $3.2\,\sigma$ ($2.7\,\sigma$) if using the test statistic technique. Thus the evidence ($> 3\sigma$) of the VBF channel is observed.

The same significance can be extracted if we scan over μ_{VBF} rather than the ratio. In this case, if we do not profile μ_{ggF} and do a simultaneous fit over the two signal strength, we can check the compatibility with the SM prediction of the ggF and VBF production. The result is presented in Fig. 5.31 [49]. The best-fit values are [49]:

$$\begin{array}{lllll} \mu_{\text{ggF}} & = 1.02 \pm 0.19 & ^{+0.22}_{-0.18} & = 1.02 & ^{+0.29}_{-0.26} \\ \mu_{\text{VBF}} & = 1.27\,^{+0.44}_{-0.40} & ^{+0.29}_{-0.21} & = 1.27 & ^{+0.53}_{-0.45}\,. \\ & (\text{stat.}) & (\text{syst.}) & & (\text{tot.}) \end{array} \tag{5.58}$$

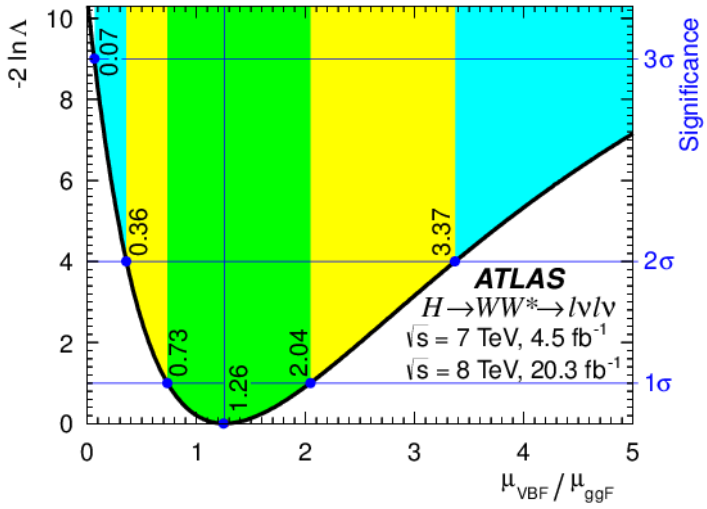


Figure 5.30: Likelihood scan as a function of $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}}$ for $m_H = 125.36\,\mathrm{GeV}$. The value of the likelihood at $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}} = 0$ gives the significance of the VBF signal at 3.2 standard deviations. The inner (middle) [outer] band shaded darker (lighter) [darker] represents the one (two) [three] standard deviation uncertainty around the central value represented by the vertical line.

5.8.3 Higgs couplings to fermions and vector bosons

By defining and fitting the signal strength μ, which is in fact a rescaling of the SM prediction, we can check the compatibility of the signal with the SM. Similarly, using a framework detailed in [61], we can define rescaling factors for the Higgs coupling to fermions and vector bosons as κ_F and κ_V, fitting on which the compatibility of the SM couplings can be checked ($\kappa_F = \kappa_V = 1$ corresponds to the SM).

The two signal strengths μ_{ggF} and μ_{VBF} can be expressed as functions of the two newly defined scale factors [61]:

$$\mu_{\mathrm{ggF}} \propto \frac{\kappa_F^2 \cdot \kappa_V^2}{(\mathcal{B}_{H\to f\bar{f}} + \mathcal{B}_{H\to gg})\kappa_F^2 + (\mathcal{B}_{H\to VV})\kappa_V^2} \tag{5.59}$$

$$\mu_{\mathrm{VBF}} \propto \frac{\kappa_V^4}{(\mathcal{B}_{H\to f\bar{f}} + \mathcal{B}_{H\to gg})\kappa_F^2 + (\mathcal{B}_{H\to VV})\kappa_V^2}\,. \tag{5.60}$$

The denominator expresses the total decay width of the Higgs in terms of its fermion decay branching ratios $\mathcal{B}_{H\to f\bar{f}}$ and $\mathcal{B}_{H\to gg}$, which takes the value of ~ 0.75, plus the vector boson decay branching ratio $\mathcal{B}_{H\to VV}$. So the likelihood scan can be done over the two scales, whose results are shown in Fig. 5.32 [49].

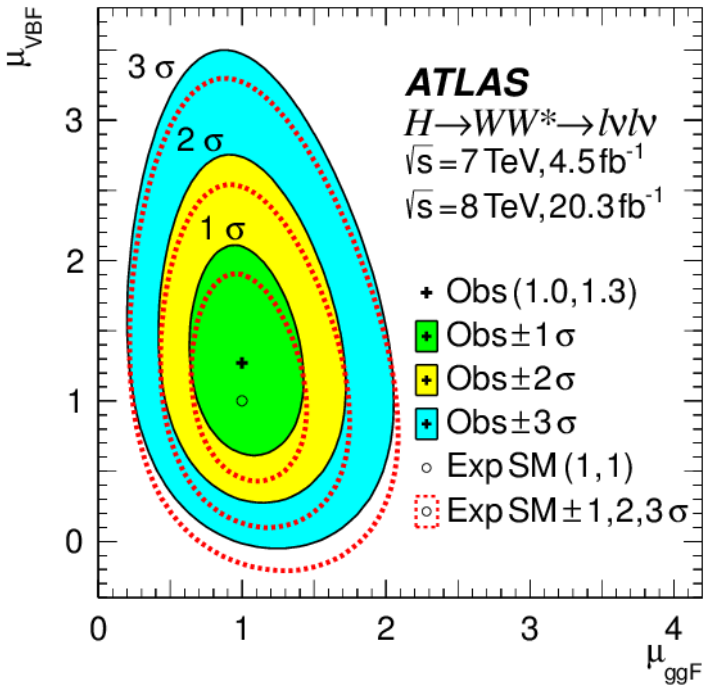


Figure 5.31: Likelihood scan as a function of μ_{ggF} and μ_{VBF}. The best-fit observed (expected SM) value is represented by the cross symbol (open circle) and its one, two, and three standard deviation contours are shown by solid lines surrounding the filled areas (dotted lines). The x- and y-axis scales are the same to visually highlight the relative sensitivity.

The low constraint power of the fit in high κ_F is easily understood, since in this case, the κ_F in the expression for μ_{ggF} cancels, which means that the ggF production mode is insensitive to κ_F. Meanwhile, the μ_{VBF} goes to zero at high κ_F which also makes the VBF mode insensitive to κ_F. To be complete, the best fit values are given below [49]:

$$\begin{array}{lllll} \kappa_F & = 0.93 \; {}^{+0.24}_{-0.18} & {}^{+0.21}_{-0.14} & = 0.93 & {}^{+0.32}_{-0.23} \\ \kappa_V & = 1.04 \; {}^{+0.07}_{-0.08} & {}^{+0.07}_{-0.08} & = 1.04 & \pm 0.11\,. \\ & \quad \text{(stat.)} & \text{(syst.)} & & \text{(tot.)} \end{array} \tag{5.61}$$

5.8.4 Cross section

To measure the central value of the cross section, it is just a matter of multiplying the signal strength μ (or μ_{ggF} or μ_{VBF}) with the corresponding SM cross section σ (or σ_{ggF} or σ_{VBF}). However, the uncertainties should be calculated carefully to exclude those related to the cross sections which do not apply to this measurement.

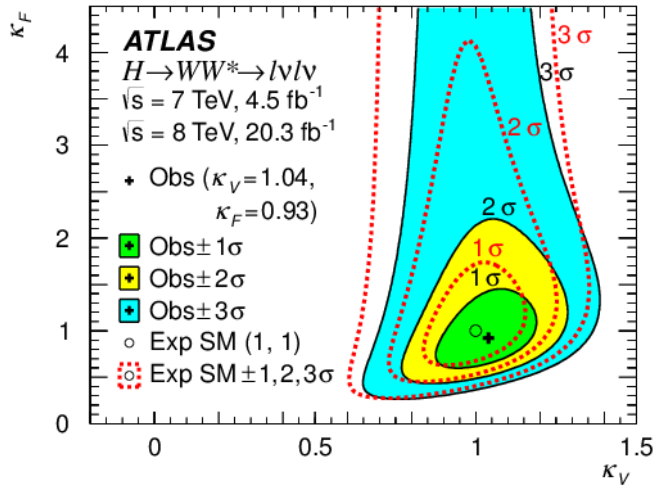


Figure 5.32: Likelihood scan as a function of κ_V and κ_F. The best-fit observed (expected SM) value is represented by the cross symbol (open circle) and its one, two, and three standard deviation contours are shown by solid lines surrounding the filled areas (dotted lines). NB. The y-axis spans a wider range than the x-axis.

The measurements are carried out for the 7 TeV ggF, 8 TeV ggF, and 8 TeV VBF signals (7 TeV VBF is not included since it suffers from large statistical uncertainty). Thus three signal strengths $\mu_{\text{ggF}}^{7\,\text{TeV}}$, $\mu_{\text{ggF}}^{8\,\text{TeV}}$, and $\mu_{\text{VBF}}^{8\,\text{TeV}}$ are introduced in the likelihood fit. The results are [49]:

$$\begin{array}{lllll} \mu_{\text{ggF}}^{7\,\text{TeV}} = 0.57 & {}^{+0.52}_{-0.51} & {}^{+0.36}_{-0.34} & {}^{+0.14}_{-0.004} \\ \mu_{\text{ggF}}^{8\,\text{TeV}} = 1.09 & \pm 0.20 & {}^{+0.19}_{-0.17} & {}^{+0.14}_{-0.09} \\ \mu_{\text{VBF}}^{8\,\text{TeV}} = 1.45 & {}^{+0.48}_{-0.44} & {}^{+0.38}_{-0.24} & {}^{+0.11}_{-0.06} \\ & \text{(stat.)} & \text{(syst.)} & \text{(sig.)} \end{array} \tag{5.62}$$

where "sig." indicates the systematics on the signal yields which should not be accounted into the cross section measurements. Multiplying them with the expected cross sections, we have the measurements:

$$\begin{array}{llllll} \sigma_{\text{ggF}}^{7\,\text{TeV}}\cdot\mathcal{B}_{H\to WW^*} = 2.0 & \pm 1.7 & {}^{+1.2}_{-1.1} & = 2.0 & {}^{+2.1}_{-2.0} & \text{pb}\,, \\ \sigma_{\text{ggF}}^{8\,\text{TeV}}\cdot\mathcal{B}_{H\to WW^*} = 4.6 & \pm 0.9 & {}^{+0.8}_{-0.7} & = 4.6 & {}^{+1.2}_{-1.1} & \text{pb}\,, \\ \sigma_{\text{VBF}}^{8\,\text{TeV}}\cdot\mathcal{B}_{H\to WW^*} = 0.51 & {}^{+0.17}_{-0.15} & {}^{+0.13}_{-0.08} & = 0.51 & {}^{+0.22}_{-0.17} & \text{pb}\,, \\ & \text{(stat.)} & \text{(syst.)} & & \text{(tot.)} & \end{array} \tag{5.63}$$

while the expected cross sections are: 3.3 ± 0.4 pb, 4.2 ± 0.5 pb, and 0.35 ± 0.02 pb, respectively.

It is also interesting to measure the fiducial cross sections σ_{fid} in a fiducial volume, since the measurements enable comparisons between the observation and theoretical predictions with minimal assumptions on the kinematics of the signal and the jet multiplicity in the event. Thus in the calculation of σ_{fid}, the uncertainties related to the kinematics acceptance should be separated from the signal strength μ, which are shown below [49]:

$$\begin{aligned} \mu^{\text{ggF}}_{0j,e\mu} &= 1.39 \quad \pm 0.27 \quad {}^{+0.21}_{-0.19} \quad {}^{+0.27}_{-0.17} \\ \mu^{\text{ggF}}_{1j,e\mu} &= 1.14 \quad {}^{+0.42}_{-0.41} \quad {}^{+0.27}_{-0.26} \quad {}^{+0.42}_{-0.17}\,. \\ &\qquad\quad (\text{stat.}) \quad (\text{syst.}) \quad (\text{sig.}) \end{aligned} \tag{5.64}$$

Only the signal strength for ggF in $e\mu$ and $n_j \leq 1$ channels are measured due to the large theoretical uncertainty for the others. The corresponding cross sections, evaluated at $m_H = 125.36\,\text{GeV}$ and using the 8 TeV data, are:

$$\begin{aligned} \sigma^{\text{ggF}}_{\text{fid,0j}} &= 27.6 \; {}^{+5.4}_{-5.3} \quad {}^{+4.1}_{-3.9} \quad = 27.6 \; {}^{+6.8}_{-6.6} \quad \text{fb}\,, \\ \sigma^{\text{ggF}}_{\text{fid,1j}} &= 8.3 \; {}^{+3.1}_{-3.0} \quad {}^{+3.1}_{-3.0} \quad = 8.3 \; {}^{+3.7}_{-3.5} \quad \text{fb}\,. \\ &\qquad\quad (\text{stat.}) \quad (\text{syst.}) \qquad\quad (\text{tot.}) \end{aligned} \tag{5.65}$$

The corresponding predictions are: $19.9 \pm 3.3\,\text{fb}$ and $7.3 \pm 1.8\,\text{fb}$, respectively.

Chapter 6

Spin of the Higgs boson

In this chapter, the study of the spin quantum numbers of the Higgs boson in the $H \to WW^* \to \ell\nu\ell\nu$ channel is presented, The 8 TeV data is used to quantify the compatibility of the data with the SM spin-0 prediction and alternative spin-2 hypotheses. In comparison with the previous analysis for the WW^* channel presented in [4], the new spin analysis focuses only on the test against spin-2 hypotheses in a new framework. Due to its special sensitivity to higher order terms of some spin-2 models, the n_j channel that I have contributed to is included in the analysis in addition to the $n_j = 0$ channel.

6.1 Introduction

For the observed Higgs-like new particle at mass point 125.36 GeV, when interpreted as the SM Higgs, good compatibility between data and theoretical prediction is reached in terms of its couplings with other particles and cross sections (inclusive and fiducial). To further confirm the new particle to be the SM Higgs boson, its spin property is studied by comparing the SM model of zero spin with models of alternative spin.

This analysis follows closely the coupling analysis introduced in the previous chapter. The same set of Monte Carlo samples as the latter are used plus additional samples dedicated for the spin analysis, The same physical object identifications and selections are performed. Event categorisation is also inherited from the coupling analysis. But only the $e\mu$ lepton flavour and $n_j \leq 1$ channels are analysed, due to limited sensitivity gain from other channels. The main background processes are similar, which are: continuum WW, top, W+jets, QCD, and $Z/\gamma^* \to \tau\tau$. Besides,

the main systematics are also the same as the coupling analysis.

The analysis technique is a bit different from the coupling analysis. BDTs are trained to discriminate events with different spins and between signal and backgrounds. Multi-dimensional BDT fit is used to extract statistical results, with the final fit supports the spin-0 model, which is the SM Higgs.

6.2 Spin-2 Higgs modelling

The spin-0 Higgs sample used in the analysis is the same one used in the coupling analysis. And only the ggF production is considered since only the $n_j \leq 1$ channels are analysed.

The spin-2 Higgs boson production considered in the analysis is modelled by the Higgs characterisation model [84], which is implemented in the MadGraph5_aMC@NLO generator. This model is actually an Effective Field Theory (EFT), which by definition is only valid up to a certain energy scale $\Lambda = 1\,\text{TeV}$. The 1 TeV choice is made to account for the fact that no new physics is observed for now at lower energy scale. The EFT approach is chosen because that it can be easily improvable by just adding higher dimensional operators in the Lagrangian. Besides, the MG5 generator include higher order calculations than the JHU which is used by the previous publication [4].

For the spin-2 Higgs boson modelling, there are a large number of choices. It is not possible to compare all of them with the spin-0 model. Instead, a specific one is chosen, which corresponds to a graviton-like tensor with minimal couplings to the SM particles [85], and has two production mechanisms of ggF and $q\bar{q}$. Its Lagrangian is defined as:

$$\mathcal{L}_2^p = \sum_{p=V,f} -\frac{1}{\Lambda}\kappa_p T_{\mu\nu}^p X_2^{\mu\nu}, \tag{6.1}$$

where $T_{\mu\nu}^p$ is the energy-momentum tensor, while $X_2^{\mu\nu}$ is the spin-2 particle field: V and f denote vector bosons (Z, W, γ and gluons) and fermions (leptons and quarks), respectively. The κ are the couplings of the Higgs-like resonance to particles [38], depending on which five different hypotheses are tested:

- Universal couplings: $\kappa_g = \kappa_q = 1$,
- $\kappa_g = 1$ and $\kappa_q = 0$, with two different Higgs-boson p_T cutoffs set at 125 and 300 GeV,

- $\kappa_g = 0.5$ and $\kappa_q = 1$, with two different Higgs-boson p_T cutoffs set at 125 and 300 GeV.

In the models with non-universal couplings, high order terms result in tails in the Higgs boson p_T spectrum. The cutoffs are introduced to prevent unitarity violation of the theory. Details about the cutoff can be found in [38]. And two different thresholds are selected to assess possible systematics. The shape of some spin-sensitive observables is affected by the tails. This feature appears in final states with at least one jet. Therefore the $n_j = 1$ channel is analysed in order to increase the sensitivity to these production modes.

6.3 Event selection

As mentioned in the introduction, the same object definitions are used as the coupling analysis. But the event selection differs due to the task of the spin analysis is to distinguish between different signal models in addition to between signals and backgrounds, while for the coupling analysis only the identification power for the latter is required.

6.3.1 Preselection

Under the same conventions or notations as in the coupling analysis, the majority of the preselection overlaps between the two analyses, with the following modification:

- Lepton flavour combination is limited to $e\mu$ or μe: large Drell-Yan backgrounds prevent significant gain of sensitivity from the $ee/\mu\mu$ channel,
- $p_T^{\text{sublead}} > 15\,\text{GeV}$: sub-leading lepton p_T threshold is modified to 15 GeV. However, the difference is negligible after considering systematic effects.

After the preselection, dedicated event cuts are designed for the $n_j = 0$ and $n_j = 1$ analyses, respectively.

6.3.2 Event selection

The $n_j = 0$ and $n_j = 1$ event selection follows that used in the coupling analysis but with looser cut thresholds. This is for the consideration of increasing spin-2 signal acceptance which is more background-like.

For the $n_j = 0$ channel, the following modifications are applied to the $e\mu/\mu e$ $n_j = 0$ cuts of the coupling analysis:

- $\Delta\phi_{\ell\ell,E_T^{miss}}$ cut is removed,
- $p_{T,\ell\ell} > 20\,\text{GeV}$: loosened from 30 GeV to 20 GeV,
- $m_{\ell\ell} < 80\,\text{GeV}$: loosened from 55 GeV to 80 GeV,
- $\Delta\phi_{\ell\ell} < 2.8$: loosened from 1.8 to 2.8,
- $p_T^H < 125$ or $300\,\text{GeV}$: this is only applied when testing the spin-2 models to the samples corresponding to non-universal couplings. It is the Higgs p_T cutoff mentioned before.

For the $n_j = 1$ channel, the following modifications are applied to the $e\mu/\mu e$ $n_j = 1$ cuts of the coupling analysis:

- $m_{\ell\ell} < 80\,\text{GeV}$: loosened from 55 GeV to 80 GeV,
- $\Delta\phi_{\ell\ell} < 2.8$: loosened from 1.8 to 2.8,
- $m_T < 150\,\text{GeV}$: the m_T cut is added to reject WW and top backgrounds and improve data/MC modelling while rejecting almost no signal,
- $p_T^H < 125$ or $300\,\text{GeV}$: the Higgs p_T cutoff.

All the signal region event selections used in the analysis are summarised in Table 6.1 [38].

The expected and observed event counts after the above selections are summarised in Table 6.2 [38]. The $n_j = 1$ category contains three parallel regions corresponding to test different spin-2 models with universal or non-universal couplings. There is no such division in $n_j = 0$ channel due to the fact that applying the p_T^H requirement in the $n_j = 0$ channel does not change substantially the event yields, while it has an effect in the $n_j = 1$ channel, as expected. Besides, from the table, it is seen that the $n_j = 0$ channel is the most sensitive one with almost three times larger event yields than in $n_j = 1$ channel. However, the $n_j = 1$ channel has its own advantage that it is sensitive to the high p_T tail of the Higgs boson, while $n_j = 0$ channel not. Some of the distributions after the above selections are shown for the SM Higgs in Figs. 6.1 and 6.2 [38]. Good data/MC agreements are observed.

Table 6.1: List of signal region selection requirements adopted for the spin analyses. The p_T^H selection requirement is applied to all samples when testing the spin-2 models with non-universal couplings.

Variable	Requirements
Preselection	
N_{leptons}	Exactly 2 with $p_T > 10\,\text{GeV}$, $e\mu$, opposite sign
$p_{\text{T}}^{\ell_1}$	$> 22\,\text{GeV}$
$p_{\text{T}}^{\ell_2}$	$> 15\,\text{GeV}$
$m_{\ell\ell}$	$> 10\,\text{GeV}$
$E_T^{\text{miss, track, jetCorr}}$	$> 20\,\text{GeV}$
0-jet selection	
$p_{T,\ell\ell}$	$> 20\,\text{GeV}$
$m_{\ell\ell}$	$< 80\,\text{GeV}$
$\Delta\phi_{\ell\ell}$	< 2.8
p_T^H	< 125 or $300\,\text{GeV}$ (*)
1-jet selection	
$b-$veto	No b-jets with $p_T > 20\,\text{GeV}$
$m_{\tau\tau}$	$< m_{\text{Z}} - 25\,\text{GeV}$
$M_{T,\text{max}}^W$	$> 50\,\text{GeV}$
$m_{\ell\ell}$	$< 80\,\text{GeV}$
$\Delta\phi_{\ell\ell}$	< 2.8
m_T	$< 150\,\text{GeV}$
p_T^H	< 125 or $300\,\text{GeV}$ (*)

Table 6.2: Expected and observed event yields in the signal regions for the $n_j = 0$ and $n_j = 1$ channels. For the dominant backgrounds, the expected yields are normalised using the control regions. The errors on the ratios of the data over the total background, N_{bkg}, only take into account the statistical uncertainties on the observed and expected yields.

Channel	N_{ggF}	N_{WW}	$N_{t\bar{t}}$	N_{Wt}	$N_{Z/\gamma^*\to\tau\tau}$	$N_{W+\text{jets}}$	N_{VV}	$N_{Z/\gamma^*\to e\mu/\mu e}$	N_{bkg}	Data	Data/N_{bkg}
$n_j = 0$ SR	218	2796	235	135	515	366	311	32	4390	4730	1.08 ± 0.02
$n_j = 1$ SR:	77	555	267	103	228	123	131	5.8	1413	1569	1.11 ± 0.03
$n_j = 1$ SR: p_T^H <300 GeV	77	553	267	103	228	123	131	5.8	1411	1567	1.11 ± 0.03
$n_j = 1$ SR: p_T^H <125 GeV	76	530	259	101	224	121	128	5.8	1367	1511	1.11 ± 0.03

6.4 Background estimation

As already mentioned, the main backgrounds considered here are the same as those appeared in the coupling analysis. Similar strategies are applied to estimate their contribution in the signal regions. For example, the W+jets and QCD back-

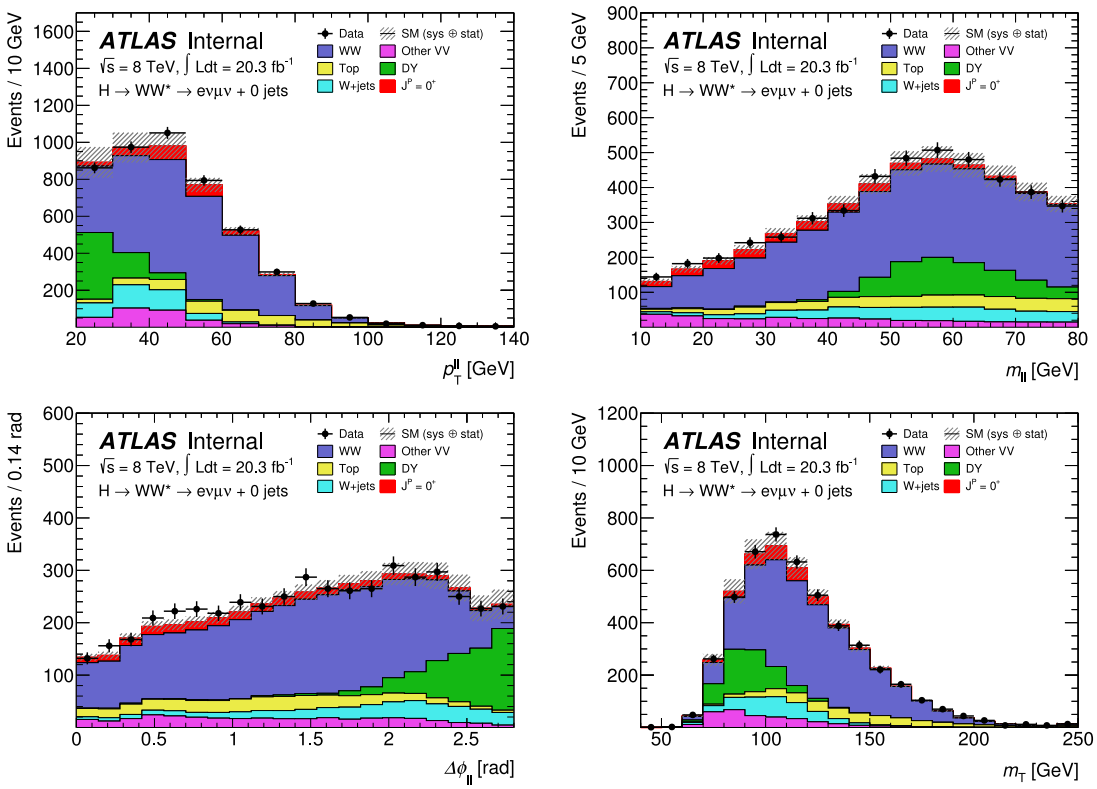


Figure 6.1: Distributions of $p_{T,\ell\ell}$, $m_{\ell\ell}$, $\Delta\,\phi_{\ell\ell}$, and m_T for the $n_j = 0$ channel. The error band includes statistical and systematic uncertainties. The signal is shown assuming a SM Higgs boson with mass $m_H = 125$ GeV. The backgrounds are normalised using control regions. The last bin in each plot includes the overflow.

grounds use exactly the same technique thus their estimation will not be repeated here. Besides, the small backgrounds of non-WW dibosons are estimated from MC predictions which are different from what were done in the coupling analysis.

For the continuum WW and the $Z/\gamma^* \to \tau\tau$ backgrounds, their estimation follows the coupling analysis, with slightly modified control region definitions and the resulting control region purities are: 69% for WW in the $n_j = 0$ channel, 43% in the $n_j = 1$ channel, 90% for Drell-Yan in the $n_j = 0$ channel, and 80% in the $n_j = 1$ channel.

For the top background in the $n_j = 0$ channel, the same JVSP method is used. As for the $n_j = 1$ channel, the old method used for the coupling analysis $n_j = 1$ top estimation is applied, which is just reversing the bveto cut to define the control region and derive the normalisation. This control region is also used to constrain the top contamination in the WW $n_j = 1$ control region which has low purity. All the control region definitions are summarised in Table 6.3 [38].

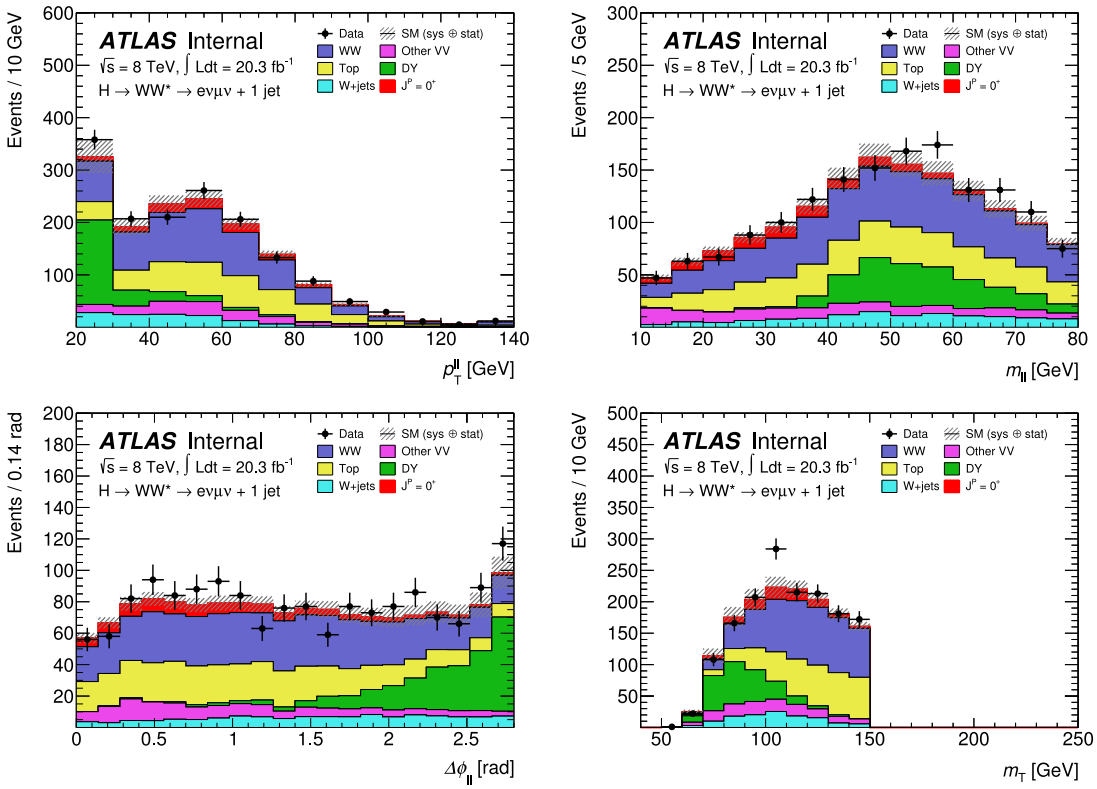


Figure 6.2: Distributions of $p_{T,\ell\ell}$, $m_{\ell\ell}$, $\Delta\,\phi_{\ell\ell}$, and m_T for the $n_j = 1$ channel. The error band includes statistical and systematic uncertainties. The signal is shown assuming a SM Higgs boson with mass $m_H = 125$ GeV. The backgrounds are normalised using control regions. The last bin in each plot includes the overflow.

Table 6.3: List of selection criteria used to define the orthogonal control regions for WW, top and $Z/\gamma^* \to \tau\tau$ backgrounds.

Control region	Selection
WW CR $n_j = 0$	Preselection, $p_{T,\ell\ell} > 20\,\text{GeV}$, $80 < m_{\ell\ell} < 150\,\text{GeV}$
WW CR $n_j = 1$	Preselection, bveto, $m_{\tau\tau} < m_Z - 25\,\text{GeV}$ $M^M_{T,max} > 50\,\text{GeV}$, $m_{\ell\ell} > 80\,\text{GeV}$
Top CR $n_j = 0$	Preselection, $\Delta\phi_{ll} < 2.8$, all jets inclusive
Top CR $n_j = 1$	At least one b-jet, $m_{\tau\tau} < m_Z - 25\,\text{GeV}$
$Z/\gamma^* \to \tau\tau$ CR $n_j = 0$	Preselection, $m_{\ell\ell} < 80\,\text{GeV}$, $\Delta\phi_{\ell\ell} > 2.8$
$Z/\gamma^* \to \tau\tau$ CR $n_j = 1$	Preselection, bveto, $M^W_{T,\text{max}} > 50\,\text{GeV}$, $m_{\ell\ell} < 80\,\text{GeV}$, $\lvert m_{\tau\tau} - m_Z\rvert < 25\,\text{GeV}$

6.5 BDT analysis

After the above selections, BDT technique is applied to analyse the data samples in the selected signal regions. Each time when comparing the compatibility with the data between the SM spin-0 model and one of the spin-2 models two BDTs are trained with a modified but similar cuts as the event selection: one BDT_0 for spin-0

and one BDT_2 for the opponent model and each BDT is trained to distinguish its corresponding signal from the various backgrounds as well as from the alternative signal. There are in total one spin-0 and five spin-2 models with their one+five BDTs in each jet channel.

Four variables are chosen as the input to the BDT training, which are the $m_{\ell\ell}$, $p_{T,\ell\ell}$, $\Delta\phi_{\ell\ell}$ and m_T. These variables are compared in terms of their discriminating shapes in Figs. 6.3 and 6.4 [38] for the $n_j = 0$ and $n_j = 1$ analyses. As an example, the two BDTs for one group of comparison between the SM Higgs and the universal coupling spin-2 Higgs are shown in Fig. 6.5 [38].

It should be point out that the input variables for the BDT training could be different in principle for the $n_j = 0$ and $n_j = 1$ channels. Studies that I have performed show that including other variables could improve slightly the BDT performance. However once the potentially large systematic uncertainty of these variables is taken into account, the final gain is found to be negligible.

6.6 Systematics

The experimental systematics inputs are the same as those used in the coupling analysis. And according to their impact on the final results, the most important ones are the jet energy scale and resolution, and the b-tagging efficiency followed by the lepton resolution and identification, trigger efficiency and missing transverse momentum. These uncertainties are summarised again here for a reminder in Table 6.4 [38].

The theoretical uncertainties on the background normalisation are estimated by studying the variations of the extrapolation factors in similar ways as in the coupling analysis for the standard systematics sources (QCD scale, PDF, UE/PS, and generator) as well as some special background-dependent sources (EW correction for WW and p_T^{Z/γ^*} reweighting). The results are summarised in Table 6.5 [38]. For the JVSP method of the $n_j = 0$ top estimation, the theoretical uncertainties are taken from the coupling analysis.

Theoretical uncertainties on the BDT shapes are also studied for the most important backgrounds. For the WW background, it is found that only the UE/PS and generator uncertainties are significantly affecting the BDT shapes. For the top background uncertainty evaluation in the $n_j = 1$ channel in which I have involved, the same group of truth-level samples I produced privately for the systematics study of the JVSP method for the coupling analysis are used and similar conclusion as the

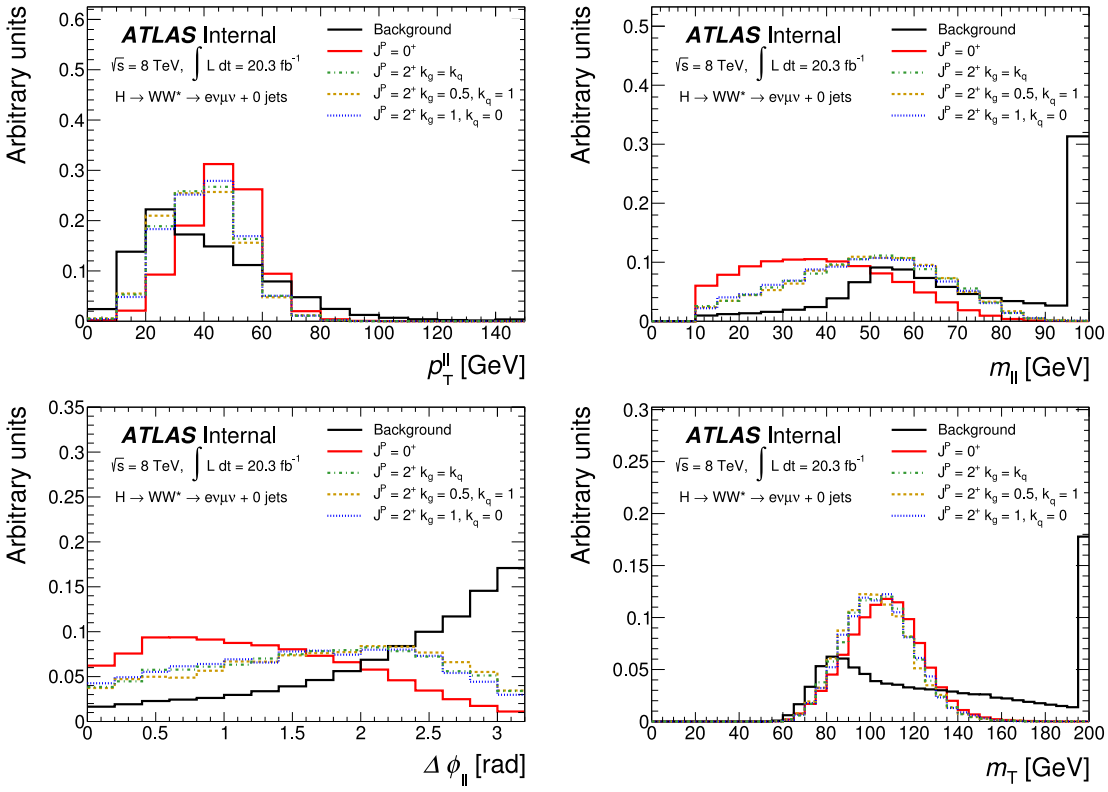


Figure 6.3: Expected normalised Higgs-boson distributions of $p_{T,\ell\ell}$, $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$ and m_T for the $e\mu$ $n_j = 0$ channel. The distributions are shown for the SM Higgs signal (solid red line) and for three spin-2 models, namely $J^P = 2^+$, $\kappa_g = 0.5$, $\kappa_q = 1$(dashed yellow line), $J^P = 2^+$, $\kappa_g = 1$, $\kappa_q = 0$ (blue dashed line) and $J^P = 2^+$, $\kappa_g = \kappa_q$ (green dashed line). The expected shapes for the sum of all backgrounds, including the data-derived W+jets background, is also shown (solid black line). The last bin in each plot includes the overflow.

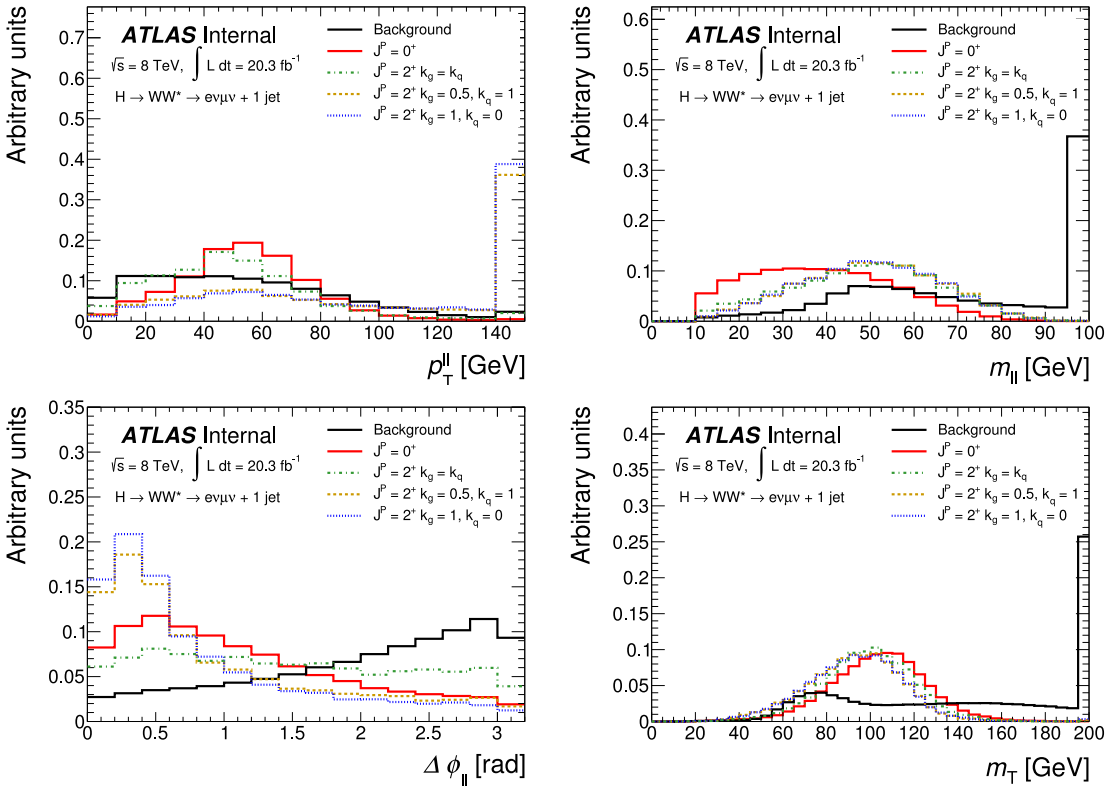


Figure 6.4: Expected normalised Higgs-boson distributions of $p_{T,\ell\ell}$, $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$ and m_T for the $e\mu$ $n_j = 1$ channel. The distributions are shown for the SM Higgs signal (solid red line) and for three spin-2 models, namely $J^P = 2^+$, $\kappa_g = 0.5$, $\kappa_q = 1$(dashed yellow line), $J^P = 2^+$, $\kappa_g = 1$, $\kappa_q = 0$ (blue dashed line) and $J^P = 2^+$, $\kappa_g = \kappa_q$ (green dashed line). The expected shapes for the sum of all backgrounds, including the data-derived W+jets background, is also shown (solid black line). The last bin in each plot includes the overflow.

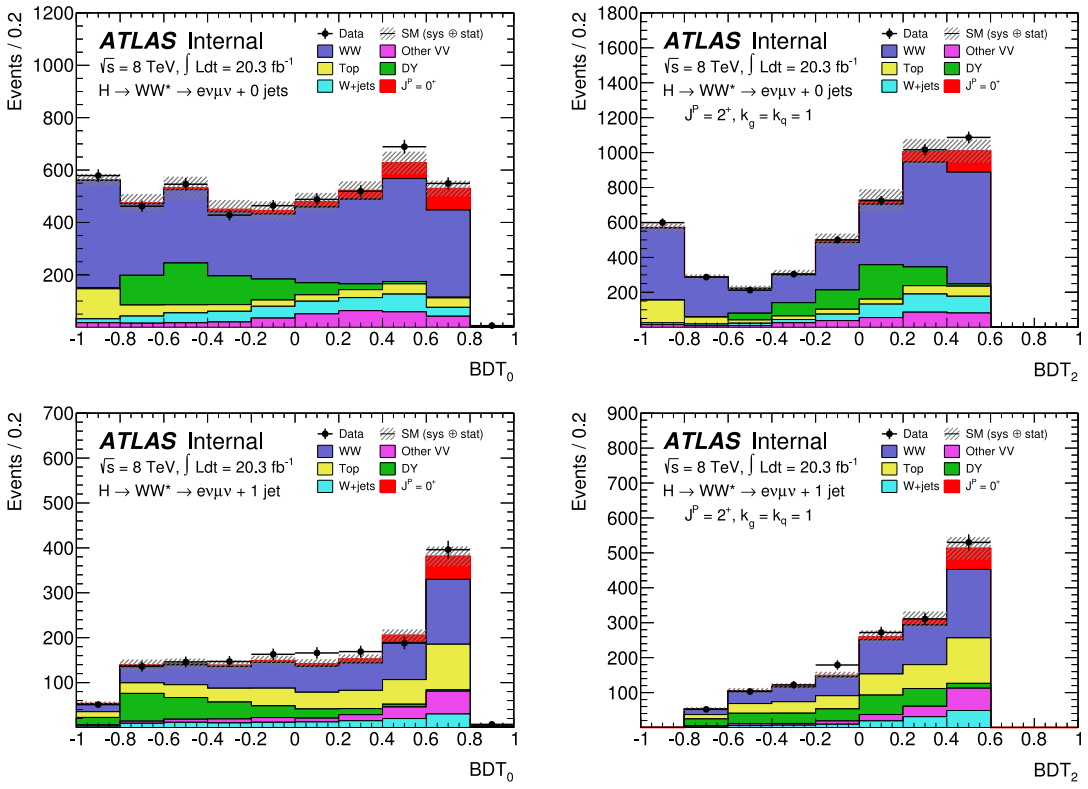


Figure 6.5: BDT_0 and BDT_2 output distributions in the signal region for spin-2 model with universal couplings. The signal is shown for the SM Higgs-boson hypothesis with $m_H = 125$ GeV. The background yields are corrected with the normalisation factors determined in the control regions.

case of WW is reached that except a 1.2% UE/PS uncertainty and 2% generator uncertainty, other sources have negligible impact.

In summary, in Table 6.6 [38], those most important systematics are ranked for a statistical test of the universal coupling spin-2 model.

6.7 Fit procedure

Before performing the fit, the two-dimensional output BDTs for each model testing are unrolled into one-dimension distributions. Since the events mainly distribute in the left bottom and right top regions of the 2-d BDT map, there will be some grids with pretty low events. By this unrolling, those bins will be merged which stabilises the fit.

The likelihood function is constructed in the same way as the coupling analysis. However, the parameter of interest here is not the signal strength μ but the fraction

Table 6.4: Sources of experimental systematic uncertainties considered in the analysis. The source and magnitude of the uncertainties and their impact on the reconstructed objects are indicated.

Source of uncertainty	Treatment in the analysis and magnitude
Jet Energy Scale (JES)	1–7% in total as a function of jet η and p_T
Jet Energy Resolution (JER)	5 – 20% as a function of jet η and p_T Relative uncertainty on the resolution is 2 – 40%
b-tagging	b-jet identification: 1 – 7.8% decomposed in p_T bins Light-quark jet misidentification: 9 – 19% as a function of η and p_T c-quark jet misidentification: 6 – 14% as a function of p_T
Leptons	Reconstruction, identification, isolation, trigger efficiency: below 1% except for electron identification: 0.2 – 2.7% depending on η and p_T Momentum scale and resolution: $< 1\%$
Missing Transverse Momentum	Propagated jet-energy and lepton-momentum scale uncertainties Resolution (1.5–3.3 GeV) and scale variation (0.3 – 1.4 GeV)
Pile-up	The amount of pile-up events is varied by 10%
Luminosity	2.8%

of the SM Higgs events with respect to the expected signal events ε, which can only takes the discrete value of 0 (for spin-2 model) or 1 (for spin-0 model). Thus the likelihood function takes the form of $\mathcal{L} = \mathcal{L}(\varepsilon, \mu, \theta)$, where μ is the signal strength (now it is a nuisance parameter) and θ the other nuisance parameters.

The test statistic is defined as:

$$q = \ln \frac{\mathcal{L}(\varepsilon = 1, \hat{\hat{\mu}}_{\varepsilon=1}, \hat{\hat{\boldsymbol{\theta}}}_{\varepsilon=1})}{\mathcal{L}(\varepsilon = 0, \hat{\hat{\mu}}_{\varepsilon=0}, \hat{\hat{\boldsymbol{\theta}}}_{\varepsilon=0})} \,. \tag{6.2}$$

Here $\hat{\hat{\mu}}_{\varepsilon=0 \text{ or } 1}$ and $\hat{\hat{\boldsymbol{\theta}}}_{\varepsilon=1 \text{ or } 1}$ are the conditional ($\varepsilon = 0$ or 1) maximum-likelihood estimators for signal strength and other nuisance parameters. The test defined here is different from the one used in the coupling analysis which is a profiled likelihood ratio. In that case, there exists a test for every possible values of the parameter of interest, which is q_μ. Here, although there are only in total two hypotheses, they share one statistic test. From the definition of the test, if the spin-0 model is correct, it will have a large positive test value, else if the spin-2 model is correct, it will have a large negative test value.

Table 6.5: Theoretical uncertainties (in %) on the extrapolation factor for WW, top and $Z/\gamma^* \to \tau\tau$ backgrounds. "Total" refers to the sum in quadrature of all uncertainties. The relative sign between rows indicates the correlation (same-sign) or anti-correlation (opposite-sign), and columns give uncorrelated sources of uncertainty.

Category	Scale	PDF	Gen	EW	UE/PS	p_T^Z	Total
WW extrapolation factor							
SR $n_j = 0$	0.9	3.8	6.9	−0.8	−4.1	–	8.2
SR $n_j = 1$	1.2	1.9	3.3	−2.1	−3.2	–	5.3
Top background extrapolation factor							
SR $n_j = 1$	−0.8	−1.4	1.9	–	2.4	–	3.5
WW CR $n_j = 1$	0.6	0.3	−2.4	–	2.0	–	3.2
$Z/\gamma^* \to \tau\tau$ extrapolation factor							
SR $n_j = 0$	−7.1	1.3	–	–	−6.5	19	21.3
SR $n_j = 1$	6.6	0.66	–	–	−4.2	–	7.9
WW CR $n_j = 0$	−11.4	1.7	–	–	−8.3	16	21.4
WW CR $n_j = 1$	−5.6	2.2	–	–	−4.8	–	7.7

Table 6.6: From top to bottom, systematic uncertainties given in % with the largest impact on the spin-2 universal couplings test. For the exact meaning of the different uncertainties related to the fake factors, refer to the W+jets section in the previous chapter of the coupling analysis.

Spin-2 universal coupling	
WW generator:	2.6
p_T^{Z/γ^*} reweighting:	1.2
Fake factor (elec. stats):	1.1
Fake factor (elec. flavour):	1.0
WW UE/PS:	0.86
Fake factor (muon stats):	0.81
$Z/\gamma^* \to \tau\tau$ generator:	0.76
Fake factor (muon flavour):	0.75
Fake factor (elec. other):	0.67

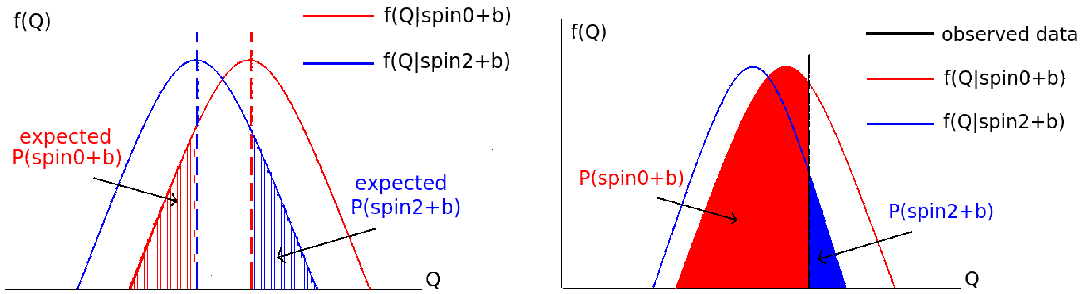


Figure 6.6: Left a) The expected $p_{\text{exp}}(2^+)$ (blue dashed area) and the expected $p_{\text{exp}}(0^+)$ (red dashed area). Right b) The observed $p_{\text{obs}}(2^+)$ (blue area) and the expected $p_{\text{obs}}(0^+)$ (red area).

The expected p-value for quantifying the compatibility with the spin-2 hypothesis $p_{\text{exp}}(2^+)$ is defined as the probability P of finding a test value q equally or less compatible with this model, which corresponds to the integral of the expected test $q_{\text{exp}(0+)}$ (the median of the sampling distribution fed with spin-0 pseudo-data) till infinite, as shown by the blue dashed area in Fig. 6.6 [86]:

$$p_{\text{exp}}(2^+) = P(q \geq q_{\text{exp}(0+)}|\epsilon = 0) = \int_{q_{\text{exp}(0+)}}^{\infty} f(q|\epsilon = 0)\, dq\,. \tag{6.3}$$

The expected p-value for compatibility with the SM hypothesis $p_{exp}(0^+)$ is defined as the probability of finding a test value q equally or less compatible with this model, which corresponds to the integral from the expected test $q_{\text{exp}(2+)}$ (the median of the sampling distribution fed with spin-2 pseudo-data) to 0, as shown by the red dashed area in Fig. 6.6(a):

$$p_{\text{exp}}(0^+) = P(q \geq q_{\text{exp}(2+)}|\epsilon = 1) = \int_{0}^{q_{\text{exp}(2+)}} f(q|\epsilon = 1)\, dq\,. \tag{6.4}$$

Besides, the confidence level for excluding an alternative theory and in favour of the SM Higgs is defined by the CLs method [87]:

$$\text{CL}_s = \frac{p^{2^+}_{\text{obs}}}{1 - p^{0^+}_{\text{obs}}}\,, \tag{6.5}$$

It means that the rejecting power of the alternative theory is normalised to the compatibility of the SM.

6.8 Results

After the fit, The expected and observed CLs values for each spin-2 models are summarised in Table 6.7 [38]. As can be seen, the SM Higgs is favoured in all tests. The other spin-2 models are excluded at a 84.5% C.L. for the universal couplings and 92.5% to 99.4% C.L. for the non-universal couplings.

Table 6.7: Summary of expected and observed sensitivities for various spin-2 models to the SM Higgs-boson hypothesis. The expected and observed p-values and the observed 1-CLs value are shown for each of them. The results are computed taking into account systematic uncertainties, using the combined $n_j = 0$ and $n_j = 1$ signal regions for the spin analysis.

Channel	$p_{\text{exp}}^{\text{SM}}$	$p_{\text{exp}}^{\text{spin-2}}$	$p_{\text{obs}}^{\text{SM}}$	$p_{\text{obs}}^{\text{spin-2}}$	1-CLs
Spin-2 $\kappa_g = \kappa_q$					
0+1-jet	0.0389	0.0334	0.2456	0.1173	84.5%
Spin-2 $\kappa_g = 0.5$, $\kappa_q = 1$ $p_T^H < 125\,\text{GeV}$					
0+1-jet	0.0466	0.0215	0.6854	0.0074	97.8%
Spin-2 $\kappa_g = 0.5$, $\kappa_q = 1$ $p_T^H < 300\,\text{GeV}$					
0+1-jet	0.0136	0.0039	0.5237	0.0034	99.3%
Spin-2 $\kappa_g = 1$, $\kappa_q = 0$ $p_T^H < 125\,\text{GeV}$					
0+1-jet	0.0409	0.0287	0.4208	0.0436	92.5%
Spin-2 $\kappa_g = 1$, $\kappa_q = 0$ $p_T^H < 300\,\text{GeV}$					
0+1-jet	0.0157	0.0041	0.5523	0.0027	99.4%

Chapter 7

Off-shell Higgs boson

In the previous chapters, a new Higgs-like particle is observed and its properties are measured in terms of its coupling with other particles and its spin quantum number. All the measurements show the compatibility with the SM Higgs. Now we turn to study the off-shell region of the Higgs boson. The corresponding measurement can be used to constrain indirectly the Higgs width.

7.1 Introduction

The Higgs off-shell measurement is motivated by the studies in Refs. [88–90], which have shown that the Higgs production in the off-shell regions over $2m_V$ ($V = Z, W$) in the $H \to ZZ$ and $H \to WW$ is largely enhanced making the measurement of off-shell couplings possible. By assuming the Higgs boson couplings are independent of the energy scale, the off-shell coupling measurement can be interpreted as a constraint on the Higgs width.

The analysis, as with the spin analysis, closely follows the coupling analysis. The same data and MC samples are used plus off-shell measurement dedicated signal/background samples. Objects and variable definitions are also taken from the coupling analysis. But the event categorisation is different here. The analysis is inclusive in jet multiplicity rather than being divided into different jet bins. This is due to the large theoretical uncertainty it will bring when calculating the exclusive cross section for the off-shell Higgs. Thus the event selection is different except for the preselections. Besides, the analysis focuses on the $e\mu/\mu e$ lepton flavour combination for its largest sensitivity. Correspondingly, the background estimation is also different. However, there are still overlaps with the coupling analysis, e.g.

the use of data-driven W+jets and QCD. Besides, the statistical treatments also go with the coupling analysis.

Finally, the results will be interpreted differently based on the assumptions we make, especially the indirect upper limit for the Higgs width can be set based on the following formulae:

$$\sigma_{\text{on-shell}} = \frac{g_V^2 \cdot g_F^2}{\Gamma_H}\,, \quad \sigma_{\text{off-shell}} = g_V^2 \cdot g_F^2\,. \tag{7.1}$$

When the couplings g_V and g_F in the on-shell cross section expression equal to those in the off-shell, the two measurements can be combined to give constraint on the Higgs width Γ_H.

7.2 MC samples

In the off-shell region, there is an important interference between the $H \to WW$ process and the non-resonant continuum WW process, since they share the same initial and final states as can be seen from the Feynman diagrams shown in Fig. 7.1 [91]. This interference effect is not included in the MC samples used in the coupling analysis. Thus new samples need to be produced to take into account the interference. Detailed comparison and validation studies are performed using different MC generators and finally it is decided to use MCFM as our benchmark generator for generating the ggF related processes.

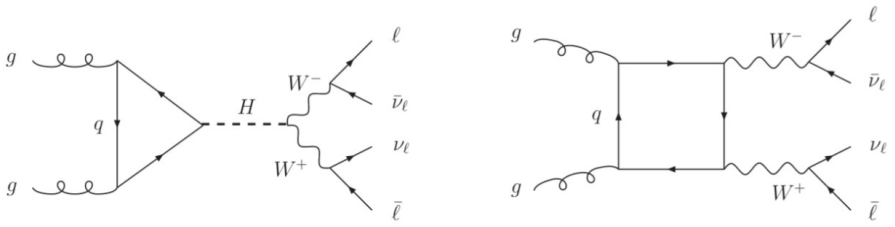


Figure 7.1: Feynman diagrams for the $gg \to H \to WW$ signal and $gg \to WW$ background.

The MCFM models in LO precision the signal and background processes, denoted as S and B. As for their interference, denoted as I, there is no dedicated sample for it, but it is contained in a sample, denoted as SBI, which includes the ggF signal, the $gg \to WW$ background, and their interference. The three samples are compared in Fig. 7.2(a) [91] as a function of the Higgs virtuality. There is

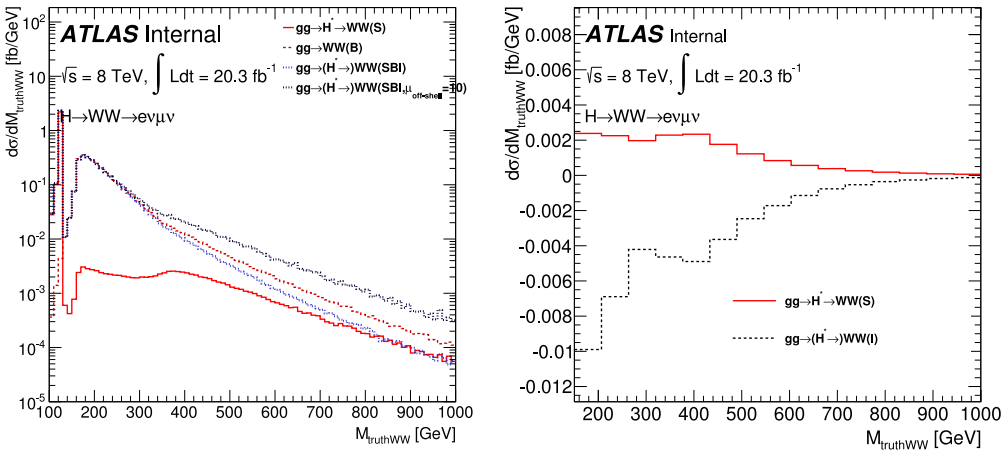


Figure 7.2: Left: Differential cross sections as a function of the invariant mass of the WW system $m_{WW}$for the signal $gg \to H \to WW$ (red solid), background $gg \to WW$ (red dash), and the former two together with their interference $gg \to (H^*) \to WW$ with default SM off-shell coupling (light blue) and ten times the SM coupling (black dash). Right: Differential cross section as a function of m_{WW} for the signal (red) and its interference with the continuum background (black).

another sample presented in the plot, whose off-shell coupling strength $\mu_{\text{off-shell}}$ is adjusted to ten times of the SM prediction. The sensitivity of the off-shell region cross section to the off-shell coupling is clearly seen. Besides, the negative feature of the interference is also observed by comparing the histograms of $gg \to WW$ with that of $gg \to (H^*) \to WW$. This is shown more clearly in Fig. 7.2(b) [91] where the interference is calculated by SBI-S-B.

Higher order QCD corrections are taken into account by reweighting the event sample with a NNLO to LO k-factor $K^{H^*}_{m_{WW}}$, shown in Fig. 7.3 [91]. This k-factor is derived from an inclusive sample, this explains why the selection of the off-shell analysis should keep the inclusiveness of the sample. For the $gg \to WW$ process, which is also in LO precision, it is corrected using the same k-factor [92] with an additional uncertainty to cover the potential bias introduced in this way. In addition, the transverse momentum of the WW systems $p_{T,WW}$ is reweighted to Sherpa which has a better description in terms of radiation jets.

The VBF production mode is also considered in the analysis. They are generated differently from those used in coupling analysis. The MadGraph+Pythia8 is used, with the inclusion of ZZ+2jets events as well as the interference between WW and ZZ channels. Four samples are generated, which correspond to the four ggF signals. Besides, the VH signal where the Higgs is on-shell and vector boson

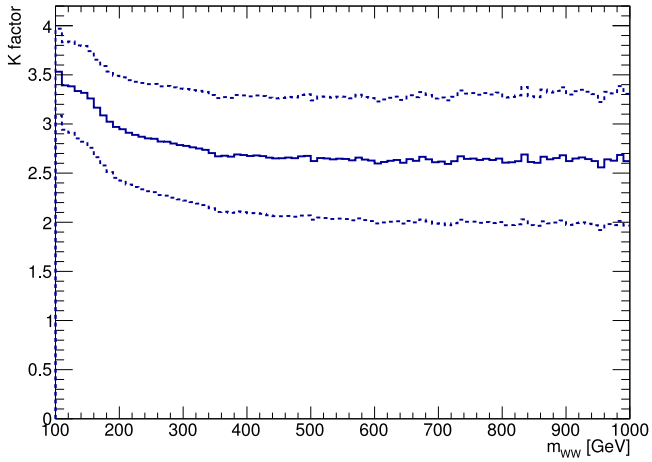


Figure 7.3: The NNLO to LO k-factor for the ggF signal process generated by MCFM as a function of the Higgs virtuality.

decays hadronicaly is also generated together with the VBF. It is removed from the VBF sample by a $|m_{WW} - m_H| > 1\,\text{GeV}$ cut and considered as background to the analysis.

7.3 Event selection

As mentioned earlier, the event preselection is exactly the same as the coupling analysis. Before introducing the signal region selection, the strategy should be mentioned:

- Minimising the on-shell contribution in the signal region: this is challenging since the Higgs mass in the WW channel cannot be fully reconstructed,
- Applying "inclusive" cuts: the selection cuts to be applied should not significantly change distributions in terms of jet multiplicity or Higgs p_T. This is due to the fact that the k factor is available only for the inclusive sample mentioned above.

To separate the off-shell Higgs signal from the on-shell one, reconstructed variables such as m_T and $m_{\ell\ell}$ are studied for their correlation with the true Higgs mass, as shown in Fig. 7.4 [91]. They are highly correlated with the Higgs mass, and the on-shell Higgs concentrate at the low value regions of both variables. This

motivates us to look at the correlation between these two variables with the results shown in Fig. 7.5 [91], which indicates again that the on-shell Higgs signal is located mostly at the left bottom corner.

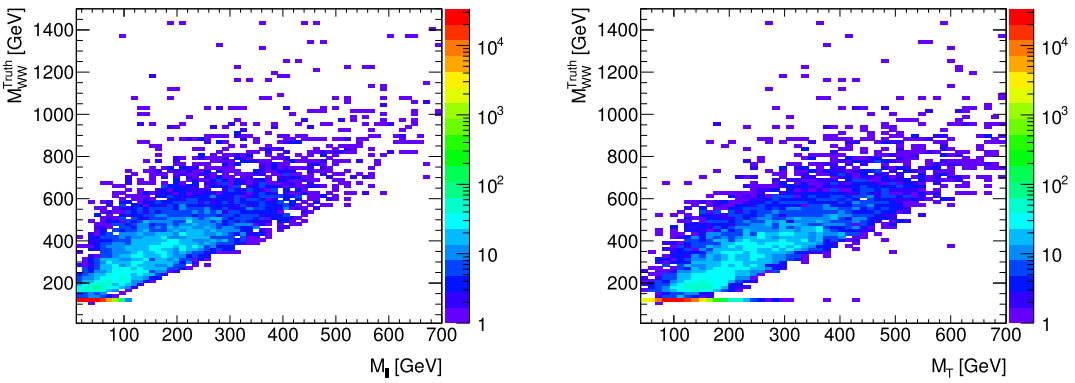


Figure 7.4: The correlation between the truth level Higgs mass and the variable $m_{\ell\ell}$ (left) and m_T (right).

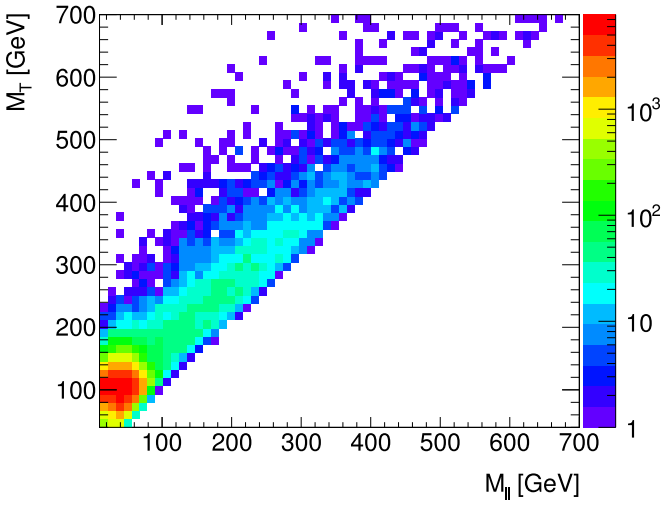


Figure 7.5: The correlation between $m_{\ell\ell}$ and m_T.

This feature inspires us to define and optimise an elliptic cut to reject the on-shell Higgs from the m_T-$m_{\ell\ell}$ plot. Thus we define a new variable:

$$R_x = \sqrt{m_{\ell\ell}^2 + (0.x \times m_T)^2}\,, \tag{7.2}$$

where x is a shape parameter of the ellipse.

To optimise the choice of x value as well as the cut to be applied on the R_x variable, a significance is defined as:

$$\text{Significance} = \frac{1}{\mu_{\text{off-shell}}} = \left(\frac{2S}{-I + \sqrt{I^2 + 4 \times 1.64 \times S\sqrt{B + (10\% \times B)^2}}} \right)^2 . \tag{7.3}$$

Here B is the total background rather than solely the continuum WW background and $\mu_{\text{off-shell}}$ is the off-shell coupling strength which is simply the solution of the following formula:

$$\mu_{\text{off-shell}} \times S + \sqrt{\mu_{\text{off-shell}}} \times I = 1.64 \times \sqrt{B}\,, \tag{7.4}$$

corresponding to an expected limit at 95% CL with S, I, B being the expected signal, interference and background yield, respectively. The 10% systematic uncertainty for the background is a rough estimation whose exact value will be estimated in later sections.

The optimal working point for the R_x variable is when $x = 0.8$ and the cut threshold is 450 GeV, which gives the best significance as shown in Fig. 7.6 [91].

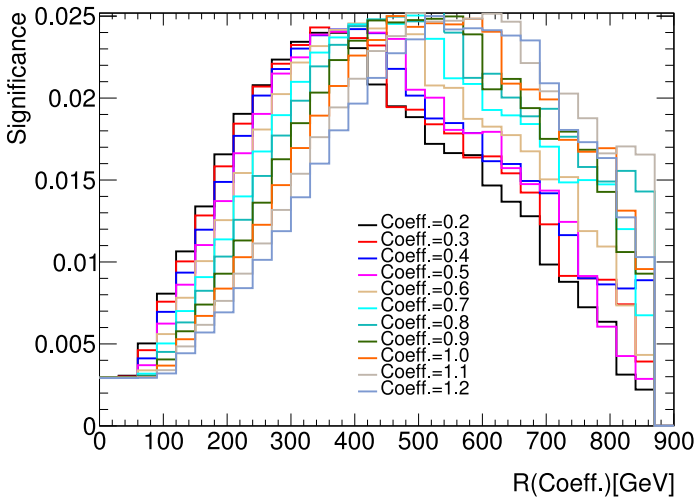


Figure 7.6: The significance scan over different x value and R_x cut.

To suppress further the WW and top backgrounds, two more cut are applied:

- b-jet veto: to suppress top,

- $\Delta\eta_{\ell\ell} < 1.2$: the cut value emerged from the same optimisation procedures presented above.

The cutflows corresponding to the above cuts as well as the preselection are shown in Table 7.1 [91] for signals and in Table 7.2 [91] for backgrounds. Some distributions after the cuts are presented in Fig. 7.7 [91] showing good data/MC agreements. Besides, the inclusiveness of the cuts are checked by looking at the evolution of the p_T and η distribution of the Higgs as well as the jet multiplicity distribution, which are shown in Fig. 7.8 [91]. From the ratio plots, it can be concluded that the inclusiveness is well kept.

Table 7.1: Cutflow table for signals and those backgrounds that interfere with signals. VH process is also shown here although treated as background.

	S(ggF)	SBI(ggF)	B($ggWW$)	S(VBF)	SBI(VBF)	SBI(VBF,r=10)	B(VBF)	S(VH)
loose preselection	650.04 ± 1.58	1474.93 ± 3.34	912.37 ± 1.86	4.14 ± 0.14	78.41 ± 1.00	91.48 ± 0.89	82.86 ± 0.62	75.71 ± 0.60
lepton p_{T}	638.39 ± 1.57	1458.41 ± 3.32	907.85 ± 1.85	4.13 ± 0.14	78.11 ± 1.00	91.27 ± 0.89	82.60 ± 0.62	74.86 ± 0.59
OS leptons	637.57 ± 1.56	1456.04 ± 3.32	905.99 ± 1.85	4.12 ± 0.14	77.84 ± 1.00	90.94 ± 0.89	82.30 ± 0.62	74.66 ± 0.59
$m_{\ell\ell} > 10$,GeV	637.57 ± 1.56	1456.04 ± 3.32	905.99 ± 1.85	4.12 ± 0.14	77.84 ± 1.00	90.94 ± 0.89	82.30 ± 0.62	74.66 ± 0.59
Z veto (for $ee, \mu\mu$)	637.57 ± 1.56	1456.04 ± 3.32	905.99 ± 1.85	4.12 ± 0.14	77.84 ± 1.00	90.94 ± 0.89	82.30 ± 0.62	74.66 ± 0.59
$E_{\mathrm{T,clj}}^{\mathrm{miss,track}} > 20$	600.87 ± 1.52	1391.87 ± 3.25	872.49 ± 1.82	3.90 ± 0.13	73.34 ± 0.97	85.80 ± 0.86	77.77 ± 0.60	67.27 ± 0.56
$R8 > 450GeV$	3.14 ± 0.10	5.72 ± 0.19	8.96 ± 0.17	0.85 ± 0.06	5.08 ± 0.26	7.68 ± 0.26	6.13 ± 0.17	0.12 ± 0.02
bveto	2.53 ± 0.09	4.84 ± 0.17	7.42 ± 0.16	0.76 ± 0.06	3.98 ± 0.22	6.21 ± 0.23	4.83 ± 0.15	0.07 ± 0.02
SR : $\Delta\eta_{\ell\ell} < 1.2$	1.54 ± 0.07	2.43 ± 0.12	3.57 ± 0.11	0.45 ± 0.04	0.98 ± 0.11	2.51 ± 0.15	1.48 ± 0.08	0.05 ± 0.02
WW CR	12.42 ± 0.20	195.73 ± 1.19	214.67 ± 0.92	0.70 ± 0.06	11.02 ± 0.37	13.21 ± 0.34	11.87 ± 0.23	2.09 ± 0.10
WW CR 0jet	6.58 ± 0.15	122.43 ± 0.98	132.82 ± 0.75	0.09 ± 0.02	2.20 ± 0.17	2.47 ± 0.15	2.27 ± 0.10	0.83 ± 0.06
WW CR inclusive	16.92 ± 0.22	234.84 ± 1.28	258.42 ± 0.98	1.76 ± 0.09	31.53 ± 0.63	36.90 ± 0.56	34.32 ± 0.39	2.92 ± 0.12
Top CR	3.37 ± 0.10	28.00 ± 0.42	32.66 ± 0.33	0.39 ± 0.04	9.38 ± 0.36	10.53 ± 0.32	9.40 ± 0.22	0.83 ± 0.06

Table 7.2: Cutflow table for backgrounds.

	$qqWW$	$WZ/ZZ/W\gamma$	Top	Z+jets	W+jets/QCD	Total Bkg.	Observed	Data/MC
loose preselection	10516.33 ± 14.56	3817.42 ± 19.35	65198.85 ± 29.64	18742.75 ± 78.70	11402.40 ± 50.62	110687.36 ± 101.12	111442	1.01 ± 0.00
lepton p_{T}	10455.97 ± 14.52	3660.35 ± 18.88	64904.28 ± 29.57	17355.72 ± 57.81	9684.15 ± 40.14	107065.16 ± 79.99	109051	1.02 ± 0.00
OS leptons	10419.81 ± 14.50	1832.56 ± 13.39	64661.29 ± 29.43	17137.20 ± 57.39	5944.07 ± 31.19	100990.77 ± 74.34	103629	1.03 ± 0.00
$m_{\ell\ell} > 10$,GeV	10419.81 ± 14.50	1832.56 ± 13.39	64661.29 ± 29.43	17137.20 ± 57.39	5902.64 ± 30.99	100949.34 ± 74.25	103629	1.03 ± 0.00
Z veto (for $ee, \mu\mu$)	10419.81 ± 14.50	1832.56 ± 13.39	64661.29 ± 29.43	17137.20 ± 57.39	5902.64 ± 30.99	100949.34 ± 74.25	103629	1.03 ± 0.00
$E_{\mathrm{T,clj}}^{\mathrm{miss,track}} > 20$	9984.15 ± 14.19	1637.05 ± 12.74	60637.25 ± 28.50	13491.55 ± 51.70	3494.86 ± 21.61	90201.92 ± 65.72	93789	1.04 ± 0.00
$R8 > 450GeV$	199.97 ± 1.96	21.29 ± 1.42	1054.31 ± 3.74	4.52 ± 1.02	20.99 ± 1.04	1316.87 ± 4.69	1262	0.96 ± 0.03
bveto	177.75 ± 1.85	17.36 ± 1.23	123.53 ± 1.24	3.54 ± 0.88	14.86 ± 0.68	349.71 ± 2.78	335	0.96 ± 0.05
SR : $\Delta\eta_{\ell\ell} < 1.2$	43.56 ± 0.87	3.60 ± 0.52	34.49 ± 0.65	1.33 ± 0.52	3.17 ± 0.30	91.31 ± 1.35	82	0.90 ± 0.10
WW CR	3255.23 ± 8.10	237.26 ± 4.49	1510.42 ± 4.25	110.13 ± 6.95	406.20 ± 4.90	5746.01 ± 13.31	5772	1.00 ± 0.01
WW CR 0jet	2245.97 ± 6.75	125.76 ± 3.37	451.47 ± 2.30	73.78 ± 5.75	253.99 ± 3.73	3286.07 ± 10.48	3352	1.02 ± 0.02
WW CR inclusive	3644.24 ± 8.54	309.28 ± 5.13	3014.16 ± 6.11	133.37 ± 9.51	483.58 ± 5.48	7879.36 ± 16.07	8007	1.02 ± 0.01
Top CR	328.04 ± 2.65	56.17 ± 2.23	12657.59 ± 12.81	12.69 ± 1.50	106.66 ± 3.82	13204.70 ± 13.89	13498	1.02 ± 0.01

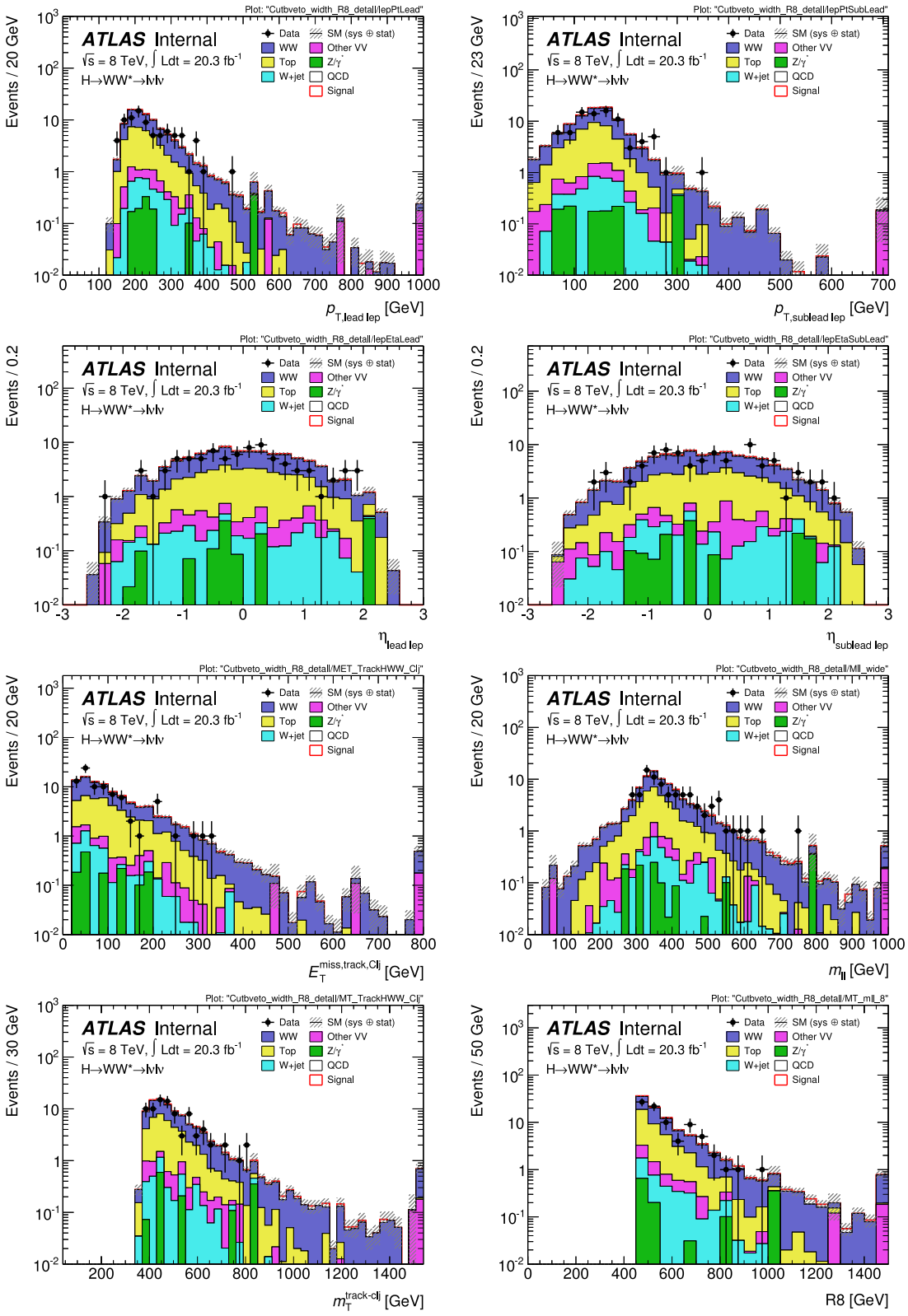


Figure 7.7: Distributions of various kinematic variables in the final selected signal region (SR).

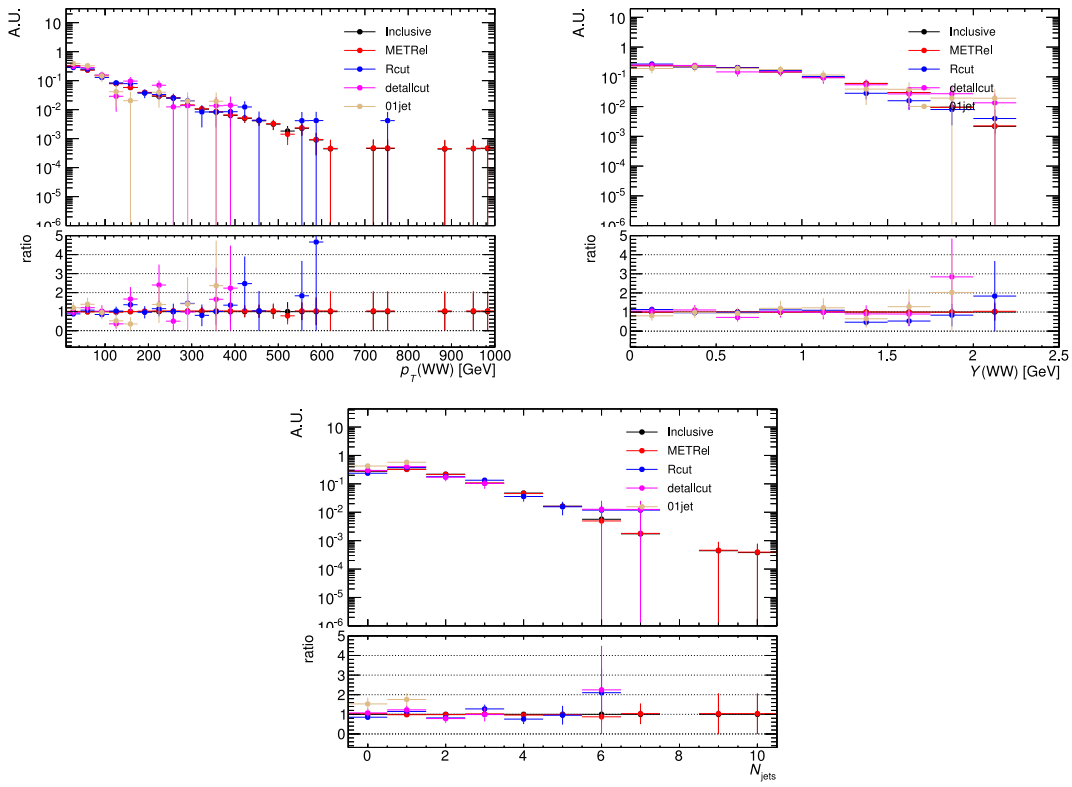


Figure 7.8: Evolution of the shape of the distributions for $p_T(WW)$ (top left), Y_{WW} (top right), and N_{jets} (bottom) after applying different selection cuts.

7.4 Background estimation

As can be seen from the cutflow table of backgrounds, the dominant backgrounds are the top and WW productions. So they are normalised using data-driven techniques. As for the others, MC simulations are used, except for the W+jets and QCD which are estimated using the fake factor method described in the coupling analysis.

The top quark background control region is defined after the preselection by:

- $R_8 > 160\,\text{GeV}$: loosened from the signal region cut at 450 GeV to increase statistics,
- $n_{b\text{-jet}} = 1$: this makes the control region orthogonal to the signal region.

The purity of the top in the CR reaches 96%. Various kinematic distributions in the top CR are presented in Fig. 7.9 [91] showing good data/MC agreements. The derived normalisation factor is $1.018 \pm 0.009(\text{stat.})$.

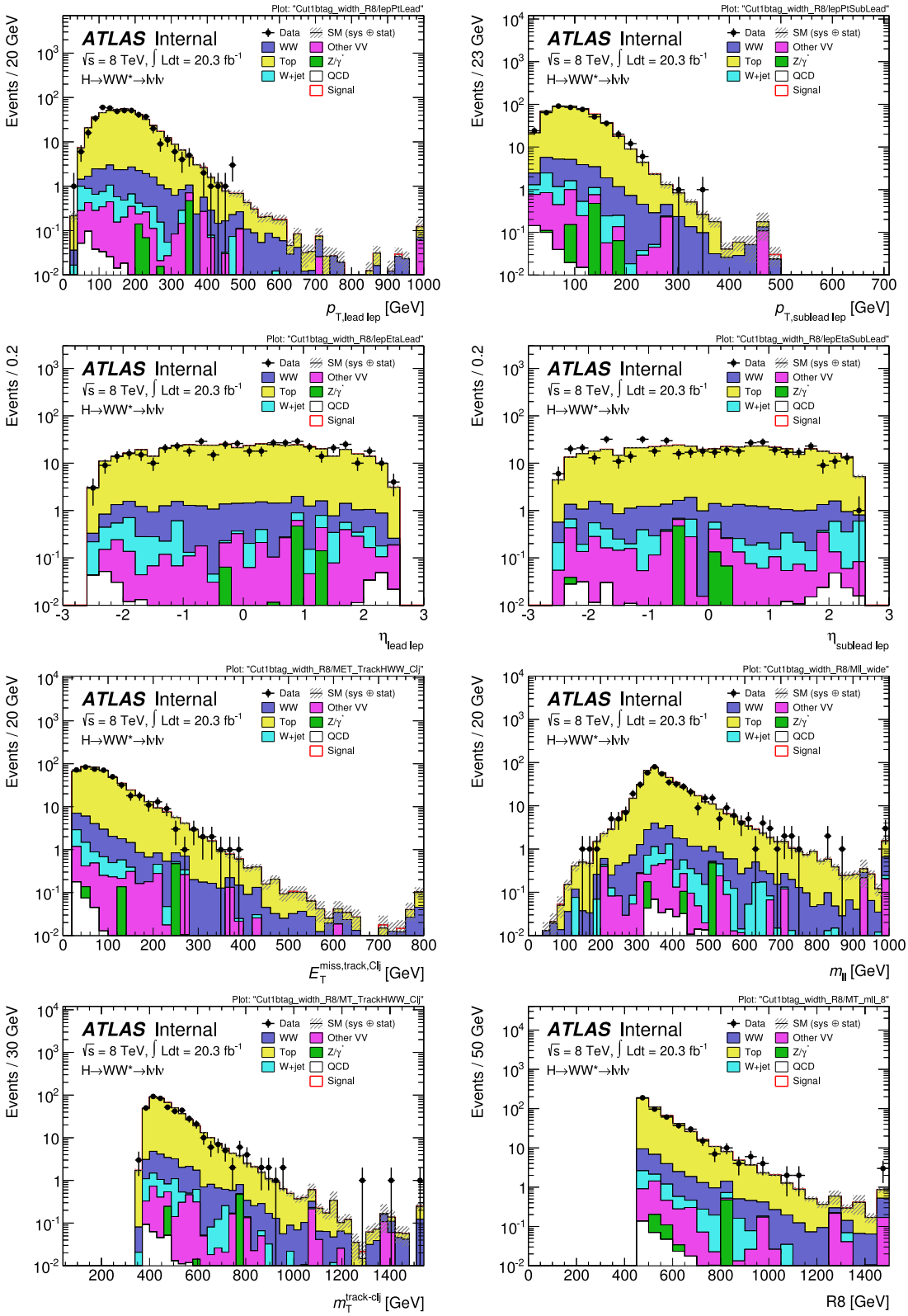


Figure 7.9: Distributions of various kinematic variables in the top CR.

The WW control region is defined following the preselection by:

- $160 < R_8 < 450\,\text{GeV}$: this makes the control region orthogonal to the signal region and away from the on-shell Higgs region,
- b-jet veto: to remove top contamination.

The corresponding kinematic distributions in the WW CR are shown in Fig. 7.10 [91]. The normalisation factor is calculated to be $0.999 \pm 0.024(\text{stat.})$, after taking into account of NNLO QCD corrections.

7.5 Systematics

The experimental systematics are estimated in the same framework as the coupling analysis, so they will not be repeated here but a conclusion is given: all systematics sources have negligible influence on the final results except for the b-tagging related uncertainties due to the dominant top background in the signal region. Thus in the following, we focus on theoretical uncertainties.

7.5.1 Top

The theoretical systematics lie in the extrapolation factor from the CR to the SR. Its variation is checked following the standard procedures of theoretical uncertainty estimation. The PDF and QCD scale uncertainties are both estimated to be 2.4%, as shown in Tables 7.3 and 7.4 [91]. As usual, the PDF uncertainty emerges from comparisons of the three PDF sets: CT10, NNPDF, and MSTW. For the QCD scale, we also varied the resummation scale up and down by a factor of two in addition to the variation on the renormalisation and factorisation scales. As for the generator and UE/PS modelling, the results are summarised in Table 7.5 [91]. The similar uncertainty for an alternative top control region definition is also shown for comparison. The one with $n_{b\text{-jet}} = 1$ is preferred as this top control region is closer to the signa region. Besides, the extrapolation from top CR to the WW CR is also studied resulting in the same PDF and scale uncertainty and 6.4% (1.9%) uncertainty on the generator (UE/PS) modelling.

7.5.2 WW

The WW here only refers to $qq \to WW$, since the $gg \to WW$ which is very much relevant to the signal due to the interference is to be studied together with

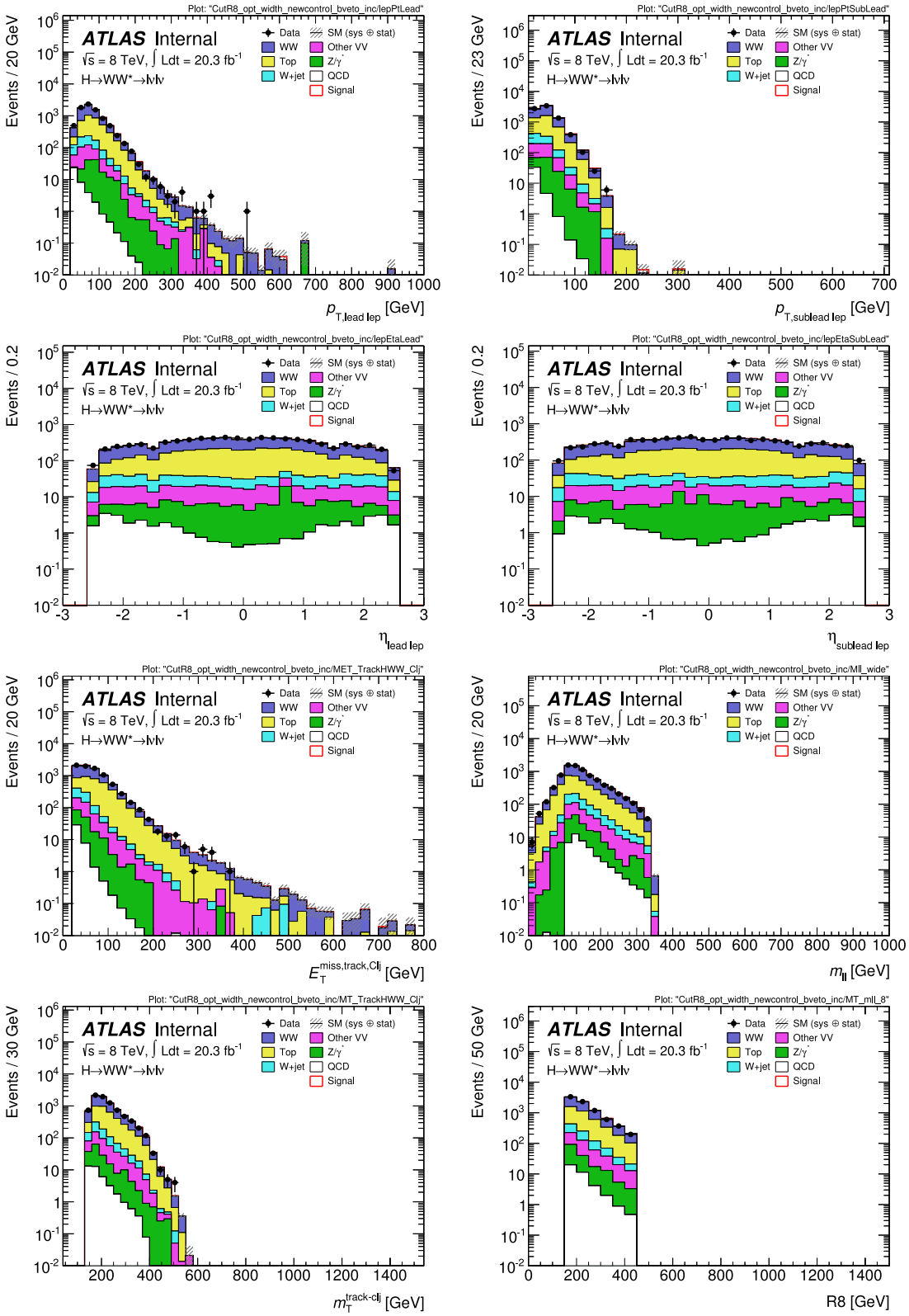


Figure 7.10: Distributions of various kinematic variables in the WW CR.

Table 7.3: Relative PDF uncertainty on the extrapolation factor (α'_{top}) from the top control region to the signal region

α'_{top}	PDF unc. (%)
CT10nlo (up)	1.3 ± 0.8
CT10nlo (down)	0.9 ± 0.8
MSTWnlo	1.0 ± 0.3
NNPDF	2.0 ± 0.3
Total	2.4

Table 7.4: Relative QCD scale uncertainty on the extrapolation factor (α'_{top}) from the top control region to the signal region

α'_{top}	QCD scale unc. (%)
Envelope (up)	2.4 ± 0.2
Envelope (down)	1.7 ± 0.2
Total	2.4

Table 7.5: Relative uncertainties due to parton shower and MC generator modelings on the extrapolation factor (α'_{top}) of from the top control region to the signal region. The systematics for an alternative top control region $n_{b\text{-jet}} \geq 1$ is also shown.

α'_{top}	PS unc. (%)	MC generator unc. (%)
Top CR: $n_{b\text{-jet}} = 1$	$5.6 \pm 0.6\%$	$1.8 \pm 0.7\%$
Top CR: $n_{b\text{-jet}} \geq 1$	$9.0 \pm 0.6\%$	$4.7 \pm 0.7\%$

the signal in later sections.

The theoretical uncertainty for the extrapolation factor from the WW CR to the SR is studied as usual. The results are summarised in Tables 7.6, 7.7, and 7.8 [91] for the generator and US/PS, PDF, and QCD systematics.

Table 7.6: Relative uncertainties due to parton shower and generator modelings on the extrapolation factor (α_{WW}) from the WW control region (CR) to the signal region.

α_{WW}	PS unc. (%)	MC generator unc. (%)
WW CR	$2.5 \pm 0.4\%$	$2.8 \pm 0.2\%$

Table 7.7: Relative PDF uncertainty on the extrapolation factor (α_{WW}) from the WW control region (CR) to the signal region.

α_{WW}	PDF unc. (%)
CT10nlo (up)	0.9 ± 1.0
CT10nlo (down)	1.0 ± 1.0
MSTWnlo	1.1 ± 0.1
NNPDF	0.8 ± 0.1
Total	1.5

Table 7.8: Relative QCD scale uncertainty on the extrapolation factor (α_{WW}) from the WW control region (CR) to the signal region.

α_{WW}	QCD scale unc. (%)
Envelope (up)	2.3 ± 0.3
Envelope (down)	1.7 ± 0.3
Total	2.3

7.5.3 Other backgrounds

As for the other backgrounds, due to their limited importance, the uncertainties are imported from the coupling analysis directly and summarised in Table 7.9 [91].

Table 7.9: Theoretical uncertainties imported from the $H \to WW^* \to \ell\nu\ell\nu$ coupling analysis. The names are as they appear in the inputs to the fit.

Source	Size	Processes
QCDscale_V	1%	Z Drell-Yan
pdf_qq	3-4%	diboson (not qqWW) and Z Drell-Yan
QCDscale_VV	5-8%	$WZZZ$, $W\gamma$, $W\gamma^*$
pdf_Wg_ACCEPT_HWW	3%	$W\gamma$
pdf_Wgs_ACCEPT_HWW	3%	$W\gamma^*$
DPI_XS	60%	Double Parton Interaction

7.5.4 Signal

For the signal processes, both their normalisations and shapes are directly taken from the MC predictions. Their uncertainties are studied in this section, which are my main contributions to this analysis.

For the PDF uncertainty of the normalisation, the method of the estimation is a bit different from the common one. It is obtained by using the function form provided by [61]:

$$\text{Reweighting factor} : w = 1 \pm 0.0066 \times \sqrt{m_{WW}/\text{GeV} - 10}\,, \tag{7.5}$$

and comparing the reweighted sample with the nominal one. The results are shown in Fig. 7.11 as a function of m_{WW} for the S, B, and SBI samples. The PDF uncertainty is similar for all three cases and is around 20% at high mass region which is our signal region.

For the evaluation of the PDF shape (acceptance) uncertainty, the PDF reweighted samples are used and three cases are considered: CT10 eigenvector set, MSTW and NNPDF. The acceptance uncertainty is quantified by the variation of the event yield in the final signal region. Since the normalisation uncertainty is already calculated and we are interested in the high mass region, each PDF varied sample is normalised in the high mass region to the nominal sample at inclusive level. The high mass region mentioned here is defined as $m_{WW} > 400\,\text{GeV}$, which is the lower bound of m_{WW} of the events after the final signal region selection. The results are summarised in the Table 7.10.

For the QCD scale systematics on the normalisation, due to the QCD NNLO to LO k-factor being applied, it is directly from the uncertainty of this k-factor. So the uncertainty is evaluated as a function of m_{WW} since the k-factor is m_{WW}

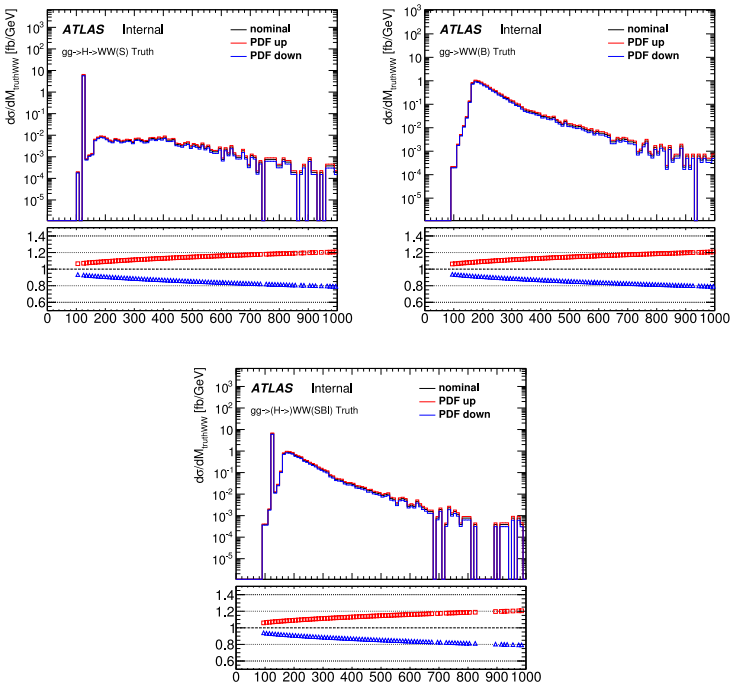


Figure 7.11: PDF uncertainty as a function of m_{WW} distributions at the truth level for the ggF Higgs, $gg \to WW$ continuum and the total of the former two plus their interference.

Table 7.10: Summary of the relative PDF acceptance uncertainty (in %) for S, B and SBI. Statistic uncertainty is omitted since the reweighted samples are heavily correlated between each other. The quadrature sum between CT10 eigenvector (EV) set and the larger one among MSTW and NNPDF is assigned as the final uncertainty.

Process	CT10 EV	MSTW	NNPDF	FINAL
S	1.5	0.3	1.8	2.3
B	1.9	0.4	2.3	3.0
SBI	2.1	0.6	2.4	3.2

dependent. The results are shown in Fig. 7.12 and the uncertainty is $\sim 20\%$ for all three samples at high mass region. Besides, as we are applying the ggF signal k-factors to the continuum WW and the SBI sample, there will be additional uncertainty introduced. To cover this bias, a universal 30% uncertainty is applied to the interference which is SBI-S-B, and a m_{WW} dependent uncertainty [-11.8%,15.9%] is assigned to both WW and the interference.

As for the generator uncertainty, it is found to be not significant, as shown in

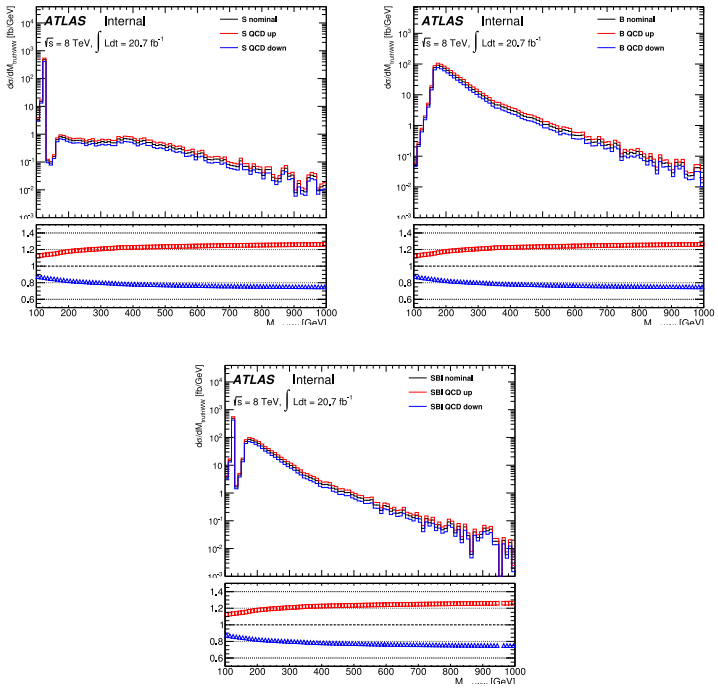


Figure 7.12: QCD normalisation uncertainty as a function of m_{WW} distributions at the truth level for the ggF Higgs, $gg \to WW$ continuum and the total of the former two plus their interference.

Fig. 7.13.

As mentioned before, the $p_{T,WW}$ shape is reweighted to the NLO prediction from Sherpa. To account for the systematics from this reweighting procedure, the QCD scale and PS uncertainties of the reweighting factors are studied, with the results summarised in Tables 7.11, 7.12, 7.13 [91] for their impacts on the signal region, WW control region and top control region. The statistical uncertainty of the reweighting factor due to limited Sherpa sample size is also considered as shown in Table 7.14 [91].

7.6 Results

The fit procedure is exactly the same as for the coupling analysis (because this is also a coupling analysis). The post-fit event yields in each regions (SR, WW CR, and top CR) for each processes are summarised in Table 7.15 [91] together with their total uncertainties. And the post-fit R_8 distributions in each regions are shown in Fig. 7.14 [91, 93].

Table 7.11: Acceptance uncertainties on the expected yield in the signal region at reconstruction level for the different processes of S, B and SBI, including the Sherpa truth $P_{T,WW}$ shape reweighting and Sherpa scale variation: The Sherpa reweighting corresponds to relative impact on MCFM from the Sherpa $P_{T,WW}$ shape reweighting; NomFdn, NomFup, RdnNom, RupNom,RupFup and RdnFdn represent different renormalisation and factorisation variation on Sherpa with the scale envelope corresponding to the largest renormalisation and factorisation variation; Qup and Qdn represent resummation variation on Sherpa with the resum envelope corresponding to the largest resummation variation. The final unc. corresponds to the larger value between half of the Sherpa reweighting uncertainty and the scale uncertainty with the two envelopes added in quadrature.

Process	S(%)	B(%)	SBI(%)
Sherpa reweighting	2.31	0.49	4.02
NomFdn	−0.14	−0.04	−0.25
NomFup	−0.11	−0.03	−0.28
RdnNom	−0.35	−0.04	−0.30
RupNom	0.02	−0.03	−0.29
RupFup	0.03	−0.03	−0.31
RdnFdn	−0.37	−0.05	−0.29
Scale envelope	0.37	0.05	0.31
Qup	0.27	−0.05	−0.13
Qdn	−0.24	−0.01	0.13
resum envelope	0.27	0.05	0.13
Final unc.	1.16	0.25	2.01

Table 7.12: Acceptance uncertainties on the expected yield in the WW control region at reconstruction level for the different processes of S, B and SBI, including the Sherpa truth $P_{T,WW}$ shape reweighting and Sherpa scale variation: The Sherpa reweighting corresponds to relative impact on MCFM from the Sherpa $P_{t,WW}$ shape reweighting; NomFdn, NomFup, RdnNom, RupNom,RupFup and RdnFdn represent different renormalisation and factorisation variation on Sherpa with the scale envelope corresponding to the largest renormalisation and factorisation variation; Qup and Qdn represent resummation variation on Sherpa with the resum envelope corresponding to the largest resummation variation. The final unc. corresponds to the larger value between half of the Sherpa reweighting uncertainty and the scale uncertainty with the two envelopes added in quadrature.

Process	S(%)	B(%)	SBI(%)
Sherpa reweighting	−6.06	−0.63	−9.11
NomFdn	0.43	0.00	−0.17
NomFup	−0.15	−0.01	−0.35
RdnNom	−0.40	−0.04	−0.85
RupNom	0.61	0.03	0.27
RupFup	0.29	0.02	0.17
RdnFdn	−0.14	−0.04	−0.78
scale envelope	0.61	0.04	0.85
Qup	−0.65	−0.02	−0.36
Qdn	1.10	0.05	0.46
resum envelope	1.10	0.05	0.46
Final unc.	3.03	0.32	4.56

Table 7.13: Acceptance uncertainties on the expected yield in the top control region at reconstruction level for the different processes of S, B and SBI, including the Sherpa truth $P_{T,WW}$ shape reweighting and Sherpa scale variation: The Sherpa reweighting corresponds to relative impact on MCFM from the Sherpa $P_{T,WW}$ shape reweighting; NomFdn, NomFup, RdnNom, RupNom,RupFup and RdnFdn represent different renormalisation and factorisation variation on Sherpa with the scale envelope corresponding to the largest renormalisation and factorisation variation; Qup and Qdn represent resummation variation on Sherpa with the resum envelope corresponding to the largest resummation variation. The final unc. corresponds to the larger value between half of the Sherpa reweighting uncertainty and the scale uncertainty with the two envelopes added in quadrature.

Process	S(%)	B(%)	SBI(%)
Sherpa reweighting	13.65	3.42	40.10
NomFdn	−0.61	−0.02	0.07
NomFup	0.39	0.01	0.37
RdnNom	0.12	0.04	0.70
RupNom	−0.39	−0.06	−0.36
RupFup	0.14	−0.05	−0.20
RdnFdn	−0.36	0.02	0.57
scale envelope	0.61	0.06	0.70
Qup	1.02	0.02	0.51
Qdn	−1.67	−0.11	−0.72
resum envelope	1.67	0.11	0.72
Final unc.	6.83	1.71	20.05

Table 7.14: Statistical uncertainty on the reweighting factor for different processes.

Process	S (%)	B (%)	SBI (%)
	1.3	1.1	1.8

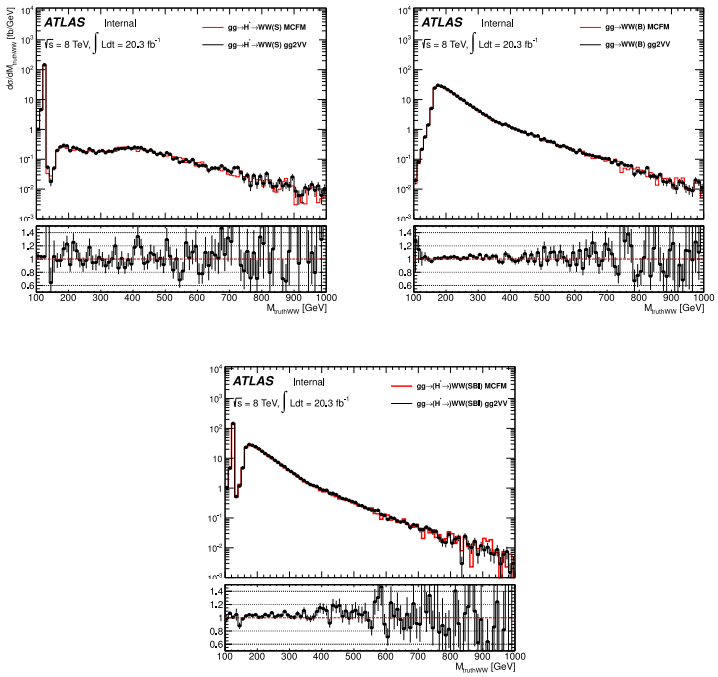


Figure 7.13: Comparison of generated m_{WW} distributions between two generators of gg2VV and MCFM for the various ggF processes: S (left), B (middle), SBI (right).

Table 7.15: The expected and observed yields, with statistical and systematic uncertainties. The SBI processes for ggF and VBF production modes are reported for both the SM off-shell coupling and 20 times the SM prediction.

Process	SR	WW CR	Top CR
$gg \to H^* \to WW$	1.5 ± 0.4	17 ± 4	3.4 ± 0.9
$gg \to WW$	3.6 ± 1.1	260 ± 60	33 ± 9
$gg \to (H^* \to)WW$	2.4 ± 1.2	240 ± 100	28 ± 12
$gg \to (H^* \to)\mathrm{WW}(\mu_{\text{off-shell}} = 20)$	22 ± 10	410 ± 170	64 ± 26
VBF $H^* \to WW$	0.42 ±0.05	1.8 ± 0.12	0.192 ± 0.019
VBF WW	1.63±0.17	37.7 ± 2.5	10.3 ± 1.1
VBF $(H^* \to)WW$	1.07 ± 0.13	34.7 ± 2.3	10.3 ± 1
VBF $(H^* \to)\mathrm{WW}(\mu_{\text{off-shell}} = 20)$	5.7 ± 0.6	52.5 ± 3.5	13.1 ±1.2
$q\bar{q} \to WW$	40 ± 5	3700 ± 400	320 ± 60
Top-quark events	35 ± 4	3070 ± 330	12940 ± 150
Other	12.2 ± 1.4	970 ± 140	194 ± 30
Total Expected (SM)	90 ± 4	8000 ± 110	13500 ± 120
Observed	82	8007	13498

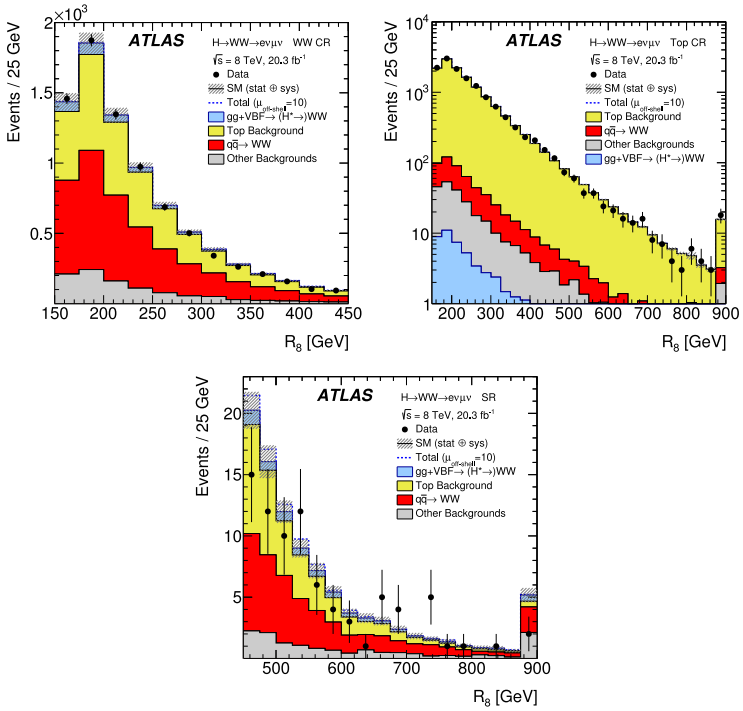


Figure 7.14: Observed distributions of R_8 for the WW control region, top control region, and the signal region including the total SM expectation and for a Higgs boson with 10 times $\mu_{\text{off-shell}}$ (dashed line).

Using the likelihood scan, the 95% C.L. upper limits on the off-shell coupling strength of $\mu_{\text{off-shell}}$ and $\mu^{\text{ggF}}_{\text{off-shell}}$ (in this case the VBF off-shell coupling takes the SM prediction) are set to be 14.6 and 18.0 with the expected ones to be 20.3 and 25.6. These results are shown in Table 7.16 and Fig. 7.15 [91, 93]. Besides, the more conservative statistical conclusion can be drawn by floating the parameter $R^B_{H^*}$ up and down by a factor of 2, which is to quantify the k-factor difference between the ggF signal and its related continuum background as well as their interference. In this case, the CLs method is used rather than the simple likelihood scan. This leads to a 95% C.L. limit on $\mu_{\text{off-shell}}$ and $\mu^{\text{ggF}}_{\text{off-shell}}$ in the range [15.6, 20.3] and [19.6, 24.7], with the expected ranges to be [18.9, 26.2] and [24.4, 32.4]. These results are summarised in Table 7.17 and shown in Fig. 7.16 [91].

These results can be interpreted as an indirect constraint on the Higgs width to be around 20 times of the SM Higgs width, namely $\sim 80\,\text{MeV}$, according to Formula 7.1 when the following assumptions are made:

- new physics modifying the off-shell couplings does not modify the SM backgrounds expectations,

Table 7.16: Observed and expected 95% C.L. upper limits from NLL scanning in the off-shell region for two different assumptions of $\mu^{\text{ggF}}_{\text{off-shell}} = \mu^{\text{VBF}}_{\text{off-shell}}$ and $\mu^{\text{VBF}}_{\text{off-shell}} = 1$, respectively.

	$-2\ln\Lambda$	Stat.	Stat.+Exp.	Stat.+Exp.+Theo.
$\mu_{\text{off-shell}}$ ($\mu^{gg}_{\text{off-shell}} = \mu^{\text{VBF}}_{\text{off-shell}}$)	Obs.	-	-	14.6
	Exp.	18.3	18.5	20.3
$\mu^{gg}_{\text{off-shell}}$ ($\mu^{\text{VBF}}_{\text{off-shell}} = 1$)	Obs.	-	-	18.0
	Exp.	21.8	22.2	25.6

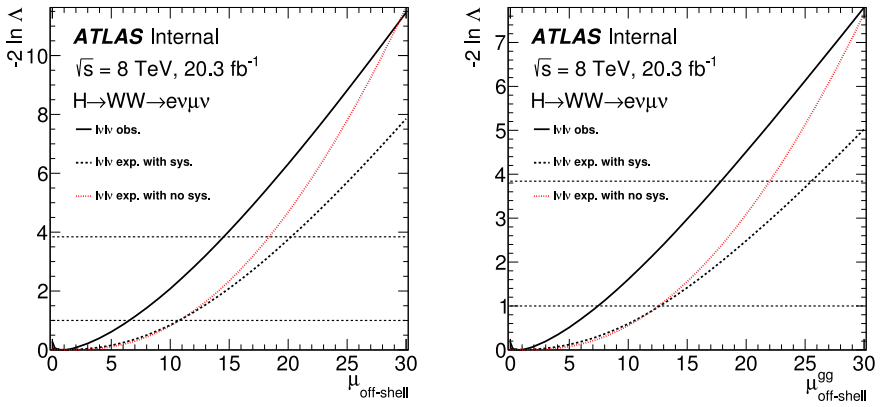


Figure 7.15: likelihood scan for $\mu_{\text{off-shell}}$ (left) and $\mu^{\text{ggF}}_{\text{off-shell}}$ (right).

- new physics does not produce sizeable signals in the search region,
- the on-shell and off-shell couplings are the same.

In summary, by counting the events in the high mass off-shell region, the upper limits of the off-shell couplings can be constrained to be around 10 to 20 times of the SM predictions [93]. When correlated with the on-shell measurements, the upper limits on the Higgs width can be extracted. These results are about two orders of magnitude more stringent than those of the direct limits derived from the $H \to \gamma\gamma$ and $H \to ZZ$ channels [7].

Table 7.17: Observed and expected 95% C.L. upper limits based on the CLs method in the off-shell region for two different assumptions of $\mu^{\text{ggF}}_{\text{off-shell}} = \mu^{\text{VBF}}_{\text{off-shell}}$ and $\mu^{\text{VBF}}_{\text{off-shell}} = 1$, respectively.

	Observed			Median expected		
$R^B_{H^*}$	0.5	**1.0**	2.0	0.5	**1.0**	2.0
$\mu_{\text{off-shell}}$	15.6	**17.2**	20.3	19.6	**21.3**	24.7
$\mu^{gg}_{\text{off-shell}}$ ($\mu^{\text{VBF}}_{\text{off-shell}} = 1$)	18.9	**21.5**	26.2	24.4	**27.2**	32.4

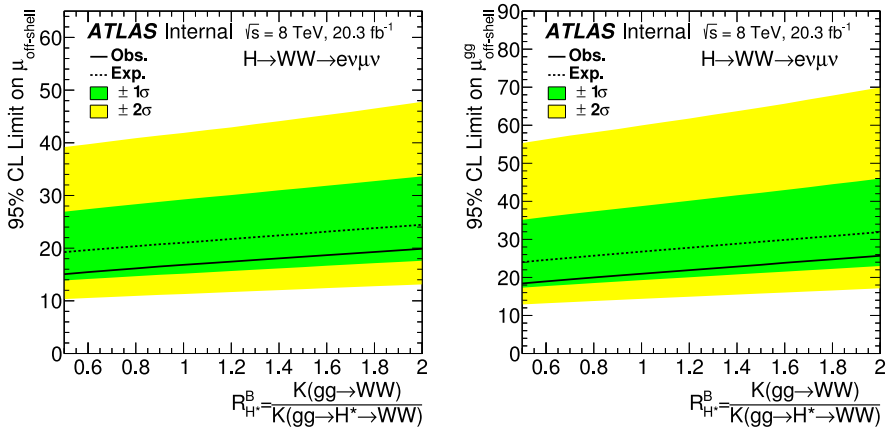


Figure 7.16: The observed and expected 95% C.L. upper limit on $\mu_{\text{off-shell}}$ (left) and $\mu^{gg}_{\text{off-shell}}$ (right) with $\mu^{\text{VBF}}_{\text{off-shell}}$ fixed to 1 as a function of R^B_{H*}.

Chapter 8

Summary and outlook

In this thesis, the main results of the Higgs research in the $H \to WW^* \to \ell\nu\ell\nu$ channel at the ATLAS are presented, which include the observation of the Higgs boson and its property measurements covering the on-shell and off-shell couplings, the spin quantum number, and a stringent constraint on the Higgs width.

The Higgs boson is observed in the $H \to WW^* \to \ell\nu\ell\nu$ channel alone with an observed (expected) significance of 6.1 (5.8) standard deviations based on an analysis using all flavour ($e\mu/\mu e$ and $ee/\mu\mu$) and jet multiplicity ($n_j = 0$, $n_j = 1$ and $n_j \geq 2$) channels and the full ATLAS Run I data corresponding to $25\,\text{fb}^{-1}$ of integrated luminosity from $\sqrt{s} = 7$ and 8 TeV pp collisions produced by the LHC at CERN [49]. This observation confirms the predicted decay of the Higgs boson to a W boson pair, at a rate consistent with that predicted by the Standard Model. In addition, evidence for VBF production in this channel is obtained with an observed (expected) significance of 3.2 (2.8) standard deviations, which further supports the Standard Model.

The new observed significance of 6.1 standard deviations is to be compared with the early results of 2.8 standard deviations in 2012 [1] and 3.8 standard deviations in 2013 [3]. Many improvements have been realised including the larger acceptance towards lower lepton p_T and reduced systematic uncertainties that I have personally contributed to.

For a Higgs boson at mass of 125.36 GeV, the on-shell coupling strengths measured independently for the ggF and VBF signals as well as for their combination

are found to be consistent with unity [49]:

$$\mu_{\mathrm{ggF}} = 1.02^{+0.19}_{-0.19}(\mathrm{stat.})^{+0.22}_{-0.18}(\mathrm{syst.}) \quad (8.1)$$

$$\mu_{\mathrm{VBF}} = 1.27^{+0.44}_{-0.40}(\mathrm{stat.})^{+0.29}_{-0.21}(\mathrm{syst.}) \quad (8.2)$$

$$\mu = 1.09^{+0.16}_{-0.15}(\mathrm{stat.})^{+0.17}_{-0.14}(\mathrm{syst.})\,. \quad (8.3)$$

Notice that the statistical precision is comparable with the systematic one for ggF and dominant for VBF. For the off-shell coupling strengths, upper limits around 20 at 95% C.L. are derived using an inclusive jet multiplicity sample selected from the $e\mu/\mu e$ channel in the 8 TeV data corresponding to an integrated luminosity of 20.3 fb^{-1} [93]. Assuming the same on-shell and off-shell coupling strengths, constraints on the upper limit of around 80 MeV on the Higgs width are achieved [93], which correspond to an improvement of about two orders of magnitude over the direct limits obtained in the $H \to \gamma\gamma, ZZ$ channels [7].

The spin property of the Higgs boson is further studied using different flavour and $n_j = 0, 1$ channels from also the 8 TeV data sample. The $n_j = 1$ channel analysis that I have contributed to is new with respect to the previous analysis [4]. The addition of this channel provides increased sensitivity to test the SM prediction of spin-0 Higgs against alternative spin-2 hypotheses. In all the tests, the SM hypothesis is favoured in data and the alternative model is disfavoured at 84.5% C.L. for the universal coupling model and excluded at 92.5% to 99.4% C.L. for the non-universal coupling models.

Within the current precision of Run I, the observed Higgs boson is found to be compatible with that of the Standard Model. However higher precision and more data are needed to tell whether it is indeed the Standard Model Higgs and whether there are other new physics beyond the Standard Model.

The RUN II operation starting in May 2015 will provide in a few years an integrated luminosity of about 100 fb^{-1} with an increased $\sqrt{s}$ from 8 TeV previously to first 13 TeV in 2015 and later 14 TeV. The Higgs production cross sections at different $\sqrt{s}$ are compared as a function of a SM-like Higgs mass in Fig. 8.1 [61]. For the SM Higgs at 125 GeV, the cross section increases by a factor of about 2.3 from 8 TeV to 13 TeV. The increase is even larger if a SM-like heavy Higgs would exist. The increase also depends on the production mode, the largest increase is for $t\bar{t}H$ reaching about 4.

The increase both in $\sqrt{s}$ and in luminosity provides a good prospect for improving precision of the Higgs property measurements in the diboson channels and

for establishing Higgs signal in the fermionic channels, as well as for searching for additional Higgs boson beyond the Standard Model.

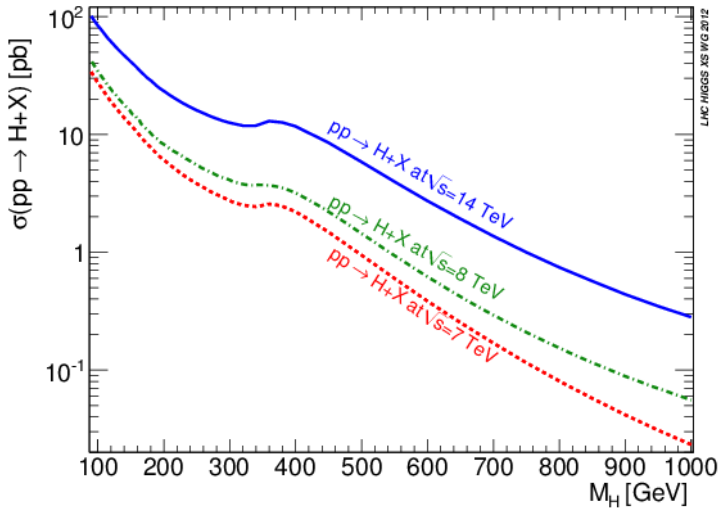


Figure 8.1: The total Higgs production cross section compared at different centre of mass energies as a function of the Higgs mass.

Appendix A

Review of data-driven methods of top background estimation

Events from both $t\bar{t}$ and single top quark processes are often one of the dominant backgrounds in the searches for new particles and measurements of Standard Model (SM) cross sections, e.g. the measurement of WW production and study for the Higgs boson in $H \to WW^* \to \ell\nu\ell\nu$ channel. So it is important to estimate their contribution precisely, usually by applying data-driven techniques. Several such methods have been proposed and used such as the JVSP method in the $n_j = 0$ channel or the in-situ b-tagging efficiency method in the $n_j = 1$ channel of the coupling analysis. However, the choice of the method is often somewhat arbitrary due to lack of comprehensive assessment on the power of each method. So in this chapter we review the various top background estimation methods and discuss the advantage or disadvantage of each and give recommendation for an optimum choice [94].

A.1 Introduction

The platform on which the review is carried out is the top background estimation in the $H \to WW^* \to \ell\nu\ell\nu$ coupling analysis. Except for those methods we used in the analysis, there are other competing ones to be discussed here. And the reason why the chosen ones are better will be explained. Besides, as already mentioned in previous chapters, the analysis is divided into three channels according to the number of selected jets in the final state: $n_j = 0$, $n_j = 1$, and $n_j \geq 2$. So the methods to be introduced in the following aim at different jet bins. However,

some of them are not confined to be applied in single bins but general methods.

A.2 Jet veto survival probability method

This method is introduced in detail in the coupling analysis chapter. Here we only give some additional comments.

The method was first applied in the cross section measurement of the SM WW process based on the 7 TeV pp collision data with an integrated luminosity of 1.02 fb^{-1} [95], the quoted relative statistical and systematic uncertainties were 15% and 20%, respectively, for the dominant different flavour channel $e\mu/\mu e$. The systematic uncertainties were dominated by a conservative estimate of the theoretical uncertainties.

For the application to the search of the SM Higgs boson in the WW channel based on the 7 TeV pp collision data corresponding to an integrated luminosity of 4.7 fb^{-1} [96], the same conservative systematic uncertainties were kept but the statistical uncertainty has been substantially reduced to 6.7%. The similar uncertainty was quoted in the Higgs discovery paper from ATLAS based on 4.7 fb^{-1} of 7 TeV and 5.3 fb^{-1} of 8 TeV data [1] and it was reevaluated with a total uncertainty of 13% in the mass and coupling measurement paper from ATLAS based on the full Run-I data [3]. In the latest $H \to WW^* \to \ell\nu\ell\nu$ coupling paper, the total systematics is further reduced to 8% dominated by experimental systematic uncertainty [49].

A.3 Template method

This method is mainly used to estimate $n_j = 0$ top background, but can also be extended to other channels.

In the template method [97], the top background estimation is performed in an extended signal region (ESR) just before applying the jet veto cut or just after the common preselection with an additional requirement of high-p_T b-jet veto. The bveto used here should be distinguished from the one used in the coupling analysis since the jet threshold is raised to 25 GeV. The ESR has the property that if one applies the jet veto cut, it becomes exactly the $n_j = 0$ signal region defined in the coupling analysis. From the ESR, we can define a top control region (CR) by requiring the appearance of a low-p_T b-jet with a transverse momentum in the range

of [20, 25] GeV. We cannot go beyond 25 GeV, since it has been vetoed already. The ESR and CR are schematically shown in Fig. A.1, together with the inclusive region and top control region used in the JVSP method for comparison.

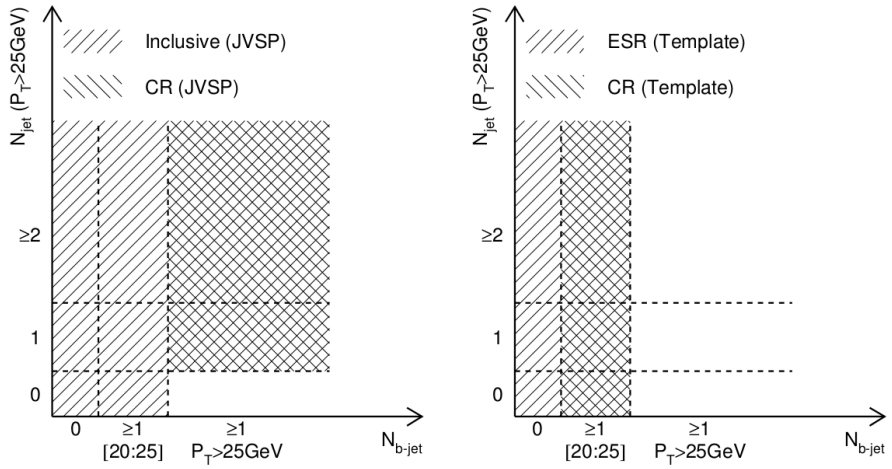


Figure A.1: Left: the inclusive and top control control (CR) regions for the JVSP method, right: the extended signal region (ESR) and top CR for the template method.

From the CR, a template of jet multiplicity distribution for the top background $T_{\text{top}}^{\text{CR}}$ can be extracted from data with the non-top contribution subtracted using MC:

$$T_{\text{top}}^{\text{CR}} = T_{\text{top}}^{\text{CR data}} - f_n \times T_{\text{non-top}}^{\text{CR MC}} , \tag{A.1}$$

where f_n corresponds to a normalisation factor for the non-top MC predicted template. This data-driven top template is extrapolated from the control region to the extended signal region using the top MC:

$$T_{\text{top}}^{\text{ESR}} = \frac{T_{\text{top}}^{\text{ESR MC}}}{T_{\text{top}}^{\text{CR MC}}} \times T_{\text{top}}^{\text{CR}} . \tag{A.2}$$

The extrapolation is done bin by bin. Then the top template $T_{\text{top}}^{\text{ESR}}$ is used to construct a template of jet multiplicity distribution for data in the ESR:

$$T_{\text{data}}^{\text{ESR}} = T_{\text{top}}^{\text{ESR}} + f_n \times T_{\text{non-top}}^{\text{ESR MC}} , \tag{A.3}$$

where the non-top part of the template is still given by the MC prediction with the same normalisation factor used in the construction of the top template in the CR $T_{\text{top}}^{\text{CR}}$. By fitting this data template to the observed data distribution in the ESR,

the scale factor f_n can be determined. Since the f_n is the only floating parameter in the top template of the ESR $T_{\text{top}}^{\text{ESR}}$, after the fit it is fully known. Thus the number of events in each bin of this template is the estimated top backgrounds in the corresponding jet bin i:

$$N_{\text{top},\,i}^{\text{est}} = (T_{\text{top}}^{\text{ESR}})_i \,. \tag{A.4}$$

In the template method, the normalisation scale factor f_n may be viewed as an effective normalisation for the various non-top background processes. It depends thus on the composition of the background processes. In the method, it is implicitly assumed that the composition is the same between the ESR and the CR. The potential difference and the corresponding systematic uncertainty was neglected. The other dominant experimental and theoretical uncertainties arise from the extrapolation factors.

For the 7 TeV pp collision data corresponding to an integrated luminosity of $4.6\,\text{fb}^{-1}$, the value of f_n was found to be 1.07 ± 0.03 [98]. For the dominant $e\mu/\mu e$ channel, the quoted relative statistical and systematic uncertainties were 26% and 15%, respectively. The large statistical uncertainty is due to the limited number of data events observed in the CR. The systematic uncertainties were dominated by the b-tagging uncertainty.

A.4 Extrapolation from control region

This is a plain idea of top background estimation which suits for all jet multiplicity channels, and usually it is realised by reversing the bveto cut in the signal region selection to construct the top control region (for those SRs without bveto, the CR is defined by requiring $n_{b\text{-jet}} \neq 0$) and then extrapolating back to the signal region. The estimated top background $N_{\text{top}}^{\text{SR, est}}$ is expressed as:

$$N_{\text{top}}^{\text{SR, est}} = N_{\text{top}}^{\text{SR, MC}} \times \frac{N_{\text{data}}^{\text{CR}} - N_{\text{non-top}}^{\text{CR, MC}}}{N_{\text{top}}^{\text{CR, MC}}} \,, \tag{A.5}$$

where the ratio term is the extrapolation factor. In fact, the extrapolation is the simplified version of the template method when the f_n parameter fit is removed.

This method is used in the top estimation of $n_j \geq 2$ channels as well as the $n_j = 1$ top estimation in spin analysis described in previous chapters. Aside its simplicity, this method usually suffers from large experimental and theoretical uncertainties due to b-tagging. For example, the quoted total uncertainties for $n_j = 1$ and $n_j \geq 2$

channels in [3] were about 30% and 40%, respectively.

A.5 In-situ b-tagging efficiency based method

This method was originally introduced from the analysis in the CMS experiment and an improved version has been applied to the $n_j = 1$ top estimation of the coupling analysis. The idea is similar to the simple extrapolation method introduced above. But the extrapolation factor here is formulated and calculated from a data-driven method rather than from MC. Considering the extrapolation from the b-tagged control region to the b-jet vetoed signal region and defining a quantity called top-tagging efficiency $\epsilon_{\text{top tag}}$ to quantify the probability of the tagger to identify top events, we have the estimated top background events $N_{\text{top}}^{\text{SR, est}}$ to be:

$$N_{\text{top}}^{\text{SR, est}} = N_{\text{top}}^{\text{CR, est}} \times \frac{1 - \epsilon_{\text{top tag}}}{\epsilon_{\text{top tag}}}, \tag{A.6}$$

where $N_{\text{top}}^{\text{CR, est}}$ is the estimated top events in the CR, which is easily calculated by subtracting the MC based non-top contribution from data, and $\frac{1-\epsilon_{\text{top tag}}}{\epsilon_{\text{top tag}}}$ is the expression for the extrapolation factor. The key is the evaluation of the efficiency $\epsilon_{\text{top tag}}$. This is done in another inclusive top control region with two jets, where we can use one jet as a probe and the other as tag to calculate a data-driven top-tagging efficiency. In ATLAS, the method is improved by adding a correction factor using top MC to take into account a possible bias of applying $\epsilon_{\text{top tag}}$ to the SR. In case where there is no bveto cut in the signal region selection, the numerator in the above formula will be removed.

In the search for the SM Higgs boson decaying in $H \to WW^* \to \ell\nu\ell\nu$ channel in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ corresponding to an integrated luminosity of $4.6\,\text{fb}^{-1}$, the quoted uncertainty by CMS was about 25% in the $n_j = 0$ channel and about 10% for the others [99]. In the VBF analysis based on the full Run-I data [100], the quoted uncertainties were about 27% and 18% at $\sqrt{s} = 7\,\text{TeV}$ and $8\,\text{TeV}$, respectively. The main uncertainty comes from the statistical uncertainty in the control sample and from the systematic uncertainties related to the measurement of the top-tagging efficiency.

A.6 Comparison between each methods

All the above methods are compared here using the data/MC samples as well as signal region definitions of the $H \to WW^* \to \ell\nu\ell\nu$ coupling analysis in terms of the normalisation factors they provide and the related uncertainties.

All four methods are compared for the $n_j = 0$ channel. The result obtained from the JVSP method has the smallest statistical uncertainty of below 3%, while those from the other methods are about a factor of 2–3 larger. This is due to the fact that those methods all rely on top control regions defined inside the $n_j = 0$ channel, which have low statistics compared to the inclusive control region used in JVSP. For the same reason, the statistical uncertainties are similar for the other methods.

The experimental uncertainty of the JVSP method is $\sim 4.6\%$, also much lower than the others because of the jet related uncertainties (which are always the dominant experimental uncertainties) cancels largely in the ratio term $\frac{P_2^{\mathrm{MC}}}{(P_1^{\mathrm{MC}})^2}$, where almost all systematic effects contribute. As an example, the experimental uncertainty on P_2^{MC} is 13.7% and on $(P_1^{\mathrm{MC}})^2$ is 12.4% to be compared with the 4.6% uncertainty when they form the ratio. For the template and extrapolation from control region methods, the experimental uncertainties are very large, which amount to 13.6% and 17.5%, respectively. This is due to the large uncertainties on the top MC based extrapolation factors as well as the relative low purity of the top control regions which introduced larger systematic effects from the non-top background subtraction. The in-situ method has a small dependence on the top MC sample since the top MC is used as a correction factor and there is also cancellation in the correction factor. However, its performance is limited by a larger experimental error attached to the non-top subtraction. The total experimental uncertainty is estimated to be 9%.

As far as theoretical systematics uncertainty is concerned, all four methods show small uncertainties, with the in-situ method giving an error of $\sim 2\%$ and the others $\sim 4\%$. In summary, the JVSP method provides a top background normalisation with the least uncertainty and is chosen to be the baseline method in the $H \to WW^* \to \ell\nu\ell\nu$ analysis.

For the top estimations in the $n_j = 1$ and $n_j \geq 2$ channels, two methods are compared: the simple control region method and the in-situ b-tagging efficiency based method. In the case of $n_j = 1$, the dominant systematics is the experimental uncertainty related to b-tagging. For the control region method, the extrapolation

factor is the ratio between a b-tagged region with a b-vetoed region, which amplifies the effects caused by the b-tagging efficiency uncertainty since the denominator and numerator vary in opposite directions. For the in-situ method, the b-tagging efficiency is measured directly in the control region from data resulting in a much smaller uncertainty and so is chosen to replace the control region method in the coupling analysis. In the case of $n_j \geq 2$, the theoretical uncertainty on the extrapolation factor dominants. The in-situ method has slightly smaller experimental systematic uncertainties.

All the above systematics studies are summarised in Table A.1 [94]. One additional type of uncertainty is added which is the non-top theoretical uncertainty. But it is always small compared to the dominant type of systematic sources in each method thus not mentioned in the above text. The experimental uncertainty is decomposed into two types based on whether its impact is on the top MC related terms (usually extrapolation factors or correction factors) or on non-top MC terms (usually used for the non-top subtraction in top control regions). The theoretical systematics is shown with detail on each type of sources including the usual sources as well as those specific to top backgrounds, e.g. the relative single top cross section uncertainty and the interference between $t\bar{t}$ and single top. Besides, the normalisation factors are also shown. In the same channel, most of them well agree within uncertainty. The worst agreement is for the NFs in the $n_j = 0$ channel between the in-situ and JVSP methods. However, the disagreement can be quantified by calculating the relative difference over the total uncertainties of the two normalisation factors, which is $\sim 1.4\sigma$ and is thus not significant.

A.7 Summary

In the search of the Higgs boson in the $WW \to \ell\nu\ell\nu$ channel at the LHC, the top background has been a dominant background source and a number of data-driven methods have been developed to determine its normalisation from data. These methods are reviewed and compared here. We have also extended the in-situ b-tagging efficiency based method by introducing a MC correction factor so that potential kinematic bias can be taken into account. The uncertainty of the top estimation varies from one method to others and may differ by several factors in terms of the total precision for a given data sample.

Furthermore, these methods can be applied to other analyses sharing the same final states both for the cross section measurements and for search for new reso-

Table A.1: Normalisation factors and their corresponding uncertainties based on the same data and MC samples used for the $H \to WW^* \to \ell\nu\ell\nu$ coupling analysis. For the theoretical uncertainties in $n_j \leq 1$, private truth-level large statistics top MC samples are used, including $t\bar{t}$ and single top. Four types of generators are used: MC@NLO+Herwig++, MC@NLO+fHerwig, POWHEG+fHerwig, and POWHEG+Pythia8. For the theoretical uncertainties of Gen/QCD/PDF in VBF, other private truth-level large statistics top MC samples are used, but including only $t\bar{t}$ process. Three types of generators are used: MC@NLO, POWHEG and Alpgen. For the rest theoretical uncertainties in VBF, the MC samples of the coupling analysis are used. For the cross talk (XTalk) uncertainty in 0/1-jet channel, a 6%/5% uncertainty is assigned to the WW/Zjets backgrounds, while in the VBF channel, an additional 26% uncertainty is assigned to the non-WW background. The labels "L", "Z", and "B" for the VBF in-situ method correspond to different choices of the dijet control regions for the evaluation of top-tagging efficiency.

Method		$n_j = 0$				$n_j = 1$		$n_j \geq 2$			
		C.R.	In-situ	Temp.	JVSP	C.R.	In-situ	C.R.	In-situ(L)	In-situ(Z)	In-situ(B)
Stat.		6.8%	7.3%	7.3%	2.2%	1.3%	2.4%	9.4%	9.5%	7.7%	9.5%
Exp.		13.6%	9.0%	17.5%	4.6%	9.5%	4.9%	7.9%	4.5%	5.4%	4.5%
Theo.	Gen.	2.4%	0.2%	3.0%	1.2%	0.2%	1.6%	14.6%	10.9%	14.6%	9.6%
	UE/PS	1.2%	0.2%	1.2%	3.2%	1.2%	0.9%	7.0%	10.2%	9.2%	12.9%
	IF	0.2%	0.2%	0.1%	0.2%	0.3%	0.7%	4.0%	3.8%	4.6%	3.8%
	QCD	1.3%	0.7%	1.6%	0.7%	0.3%	0.2%	4.3%	4.0%	4.5%	4.1%
	PDF	1.2%	1.6%	1.5%	1.3%	2.7%	1.2%	7.1%	5.4%	6.5%	5.1%
	Wt Xsec.	1.7%	0.7%	2.0%	0.5%	1.0%	1.1%	0.4%	0.3%	0.3%	0.6%
	Total	3.6%	1.9%	4.4%	3.8%	3.2%	2.6%	18.6%	16.8%	19.5%	17.8%
XTalk	WW	1.5%	1.7%	2.0%	1.1%	0.2%	0.3%	0.1%	0.2%	0.1%	0.2%
	Non-WW	—	—	—	—	—	—	0.2%	0.2%	0.2%	0.3%
	Drell-Yan	0.9%	1.1%	1.2%	1.1%	0.2%	0.3%	0.3%	0.4%	0.3%	0.5%
	Total	1.8%	2.0%	2.3%	1.5%	0.3%	0.4%	0.4%	0.5%	0.4%	0.6%
Total Uncert.		15.7%	11.9%	19.6%	6.5%	10.1%	6.1%	22.3%	19.8%	21.7%	20.7%
NF		1.33±0.21	1.33±0.16	1.25±0.25	1.08±0.07	1.03±0.10	1.06±0.06	1.47±0.33	1.47±0.29	1.39±0.30	1.48±0.31

nances or new physics. A few non exhaustive examples are listed below:

- Search for direct slepton and gaugino production in final states with two leptons and missing transverse momentum with the ATLAS detector in pp collision at $\sqrt{s} = 7\,\text{TeV}$ [101].

- Exclusive search for supersymmetry with same-flavour dilepton final states with the ATLAS detector [102].

- Search for heavy neutrinos and right-handed W bosons in events with two leptons and jets in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector [103].

- Search for heavy neutrinos and W_R bosons with right-handed couplings in a left-right symmetric model in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ [104].

- Search for narrow resonances in dilepton mass spectra in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ [105].

- Search for new phenomena in the $WW \to \nu\nu$ final state in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector [106].

In most of these analyses the top background was estimated from MC simulation and therefore can be improved (strongly recommended) by applying one of the data-driven methods discussed here.

Appendix B

Fake muon study

The following content might have little connection with the main topic of the thesis. But it is my qualification task so that I put it here. The task was to improve the fake muon rate and define a working point for the future ATLAS muon reconstruction software, so-called 3rd muon chain, to be used for analyses at Run II.

B.1 Introduction

Muon reconstruction is already introduced in Section 3.4.5, whose reconstruction efficiency reaches $\sim 90\%$. As an example, in a $t\bar{t}$ sample with all hadronic decay, the reconstruction efficiency of muon with $p_T > 10\,\text{GeV}$ is shown in Fig. B.1 with an average value of 93%. But the efficiency decreases in the central η region because there is no detector there for engineer problem. Also as p_T decreases, the efficiency reduces since the low p_T muons may have no enough momentum to traverse the muon spectrometer thus leaving only segments of tracks there or even no hits at all.

Object reconstructions at ATLAS are always accompanied with fakes. So does the muon. There are various sources for the reconstructed fake muon in ATLAS:

- Part of or all of the reconstructed combined muon tracks are created by a non-muon particle:
  - When a π/K from the interaction point decays to produce a μ, it may be reconstructed as a muon. The inner part of the track belongs to the π/K while outer part to a real μ. This is called decay-in-flight fake

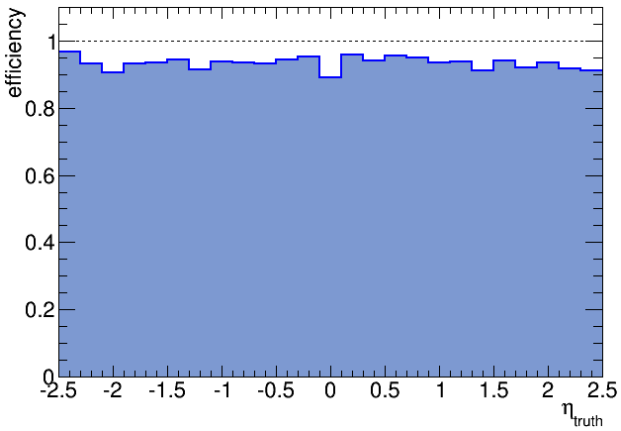


Figure B.1: Muon reconstruction efficiency as a function of η in a hadronically decaying $t\bar{t}$ sample.

muon. The decay positions (those recorded) in the R direction (in mm) of these π/K are shown in Fig. B.2(a). These points also indicate the generation positions of the fake muons which almost lie in the ID. As a comparison, the R coordinates of the true muon's generation positions are also shown in Fig. B.2(b), indicating that most true muons (prompt muons) decayed inside the beam pipe ($r = 36\,\text{mm}$).

– A proton from the interaction point survived through the calorimeter and reached the MS. This is called punch-through fake and the whole track does not correspond to any real μ.

- Real μ from pile-up events.

Generally, π/K decayed muon is the main source of fake muons.

The fake rate for muon reconstruction at ATLAS is already very small thanks to the MCP recommended muon quality cuts. But the fake rate suppression is still not enough strong. For example, in the "$B/B_s \to \mu\mu$" rare decay search, we suffer 4–5 times higher fake contamination from "$B/B_s \to hh$" (h means hadron) compared with that of the same search in CMS. In this chapter, the fake muons will be studied in terms of a possible method to suppress them.

B.2 TMVA method

Motivated by the successful application of the electron likelihood identification in the $H \to WW^* \to \ell\nu\ell\nu$ coupling analysis, we have used a TMVA technique,

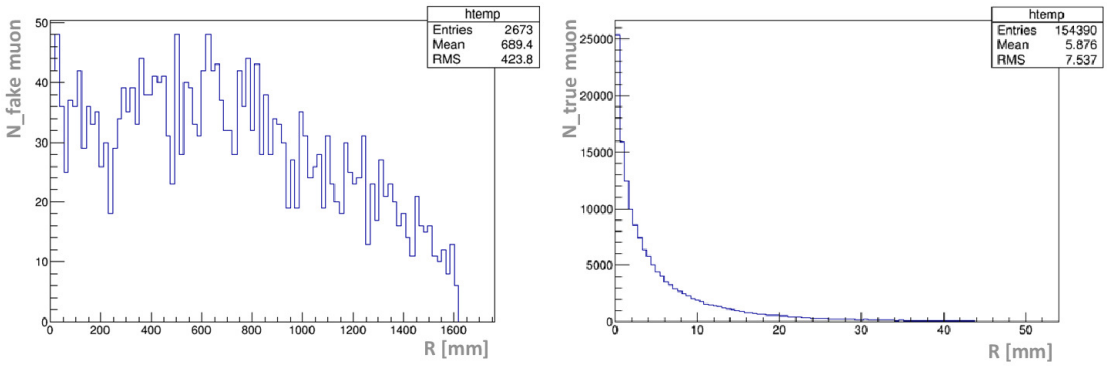


Figure B.2: Production position of fake (left) / true (right) muons measured in the transverse distance from the interaction point.

BDT, to suppress fake muons.

To search for possible discriminating variables as input to the BDT training, a $t\bar{t} \to hh$ sample is used since there are abundant fake muons in this sample. The study is based on the muon collection of the 3rd chain, to be used at Run II. Muons reconstructed with combined algorithm MUID (which means that both an ID track and an MS track must exist and the final muon track is a combined fit of the two) are used since it has the largest reconstruction efficiency. Besides, basic MCP recommended quality cuts are applied because we want to train the BDT to identify those fakes that are hard to suppress rather than those that can be easily filtered by the MCP cuts.

The study starts from making true/fake muon samples by picking up muons from the above sample and categorising them into different classes based on the truth particle type they are matched to. The matching algorithm is a cone algorithm. The muons are classified into six types: real muon, pion, kaon, proton, other hadrons, and unknown particles. The last class mainly refers to the pile-up muon which has no truth information stored. Various distributions of these classes are compared in terms of their shapes in order to search for input variables for the BDT.

After having investigated a large number of variables, the followings are selected:

- Momentum balance significance: the measured momentum difference between the ID track and the MS track, normalised to the measurement error. For non-muon particles, when they pass the calorimeter, they lose more energy than the muon does which results in large value for this quantity. The distri-

bution of this momentum balance significance is shown in Fig. B.3. The true muon is in red histogram while the fakes are in color-filled histograms. As can be seen, the true muon has a distinct shape with respect to the others.

- Scattering curvature significance: it reflects how large a particle is being scattered, normalised to the measurement error. For the decay-in-flight-fake muons, there is a dis-continuity somewhere in the reconstructed track, which results in large scatter significance. The distribution of the quantity is shown in Fig. B.4. Different shapes between true and fake muons are observed.

- $\frac{\chi^2}{\text{n.d.o.f}}$: which is the χ^2 to quantify the matching quality between the ID track and the MS track, normalised to the number of degree of freedom (n.d.o.f) of the fit. For the decay-in-flight fake muons, since the whole track is composed of two different types of particles, the matching quality must be worse than that of true muons, as can be seen in Fig. B.5.

- Qoverp ratio between the MS and ID tracks: Qoverp means q/p, which is the ratio between the track charge and track momentum, similar to the momentum balance significance, as shown in Fig. B.6.

- Number of precision layer holes: the hole is defined as the missing measurement in a layer of the MS detector when it is expected to have one. The true muon tends to have no holes as shown in Fig. B.7.

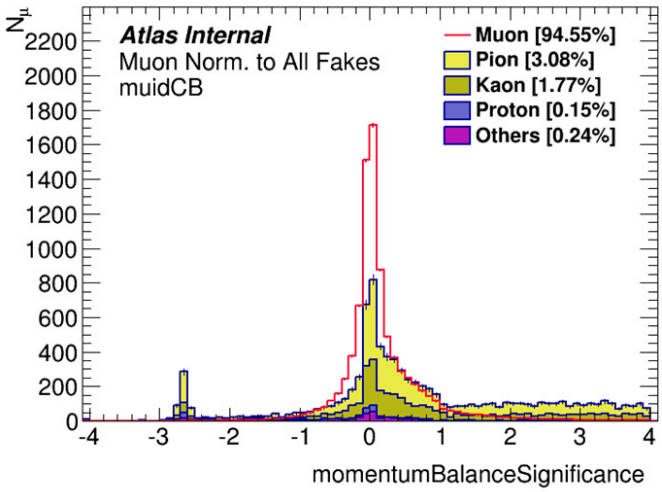


Figure B.3: Comparison of the momentum balance significance for true muons and various fake muons. All distributions are normalised to the same number.

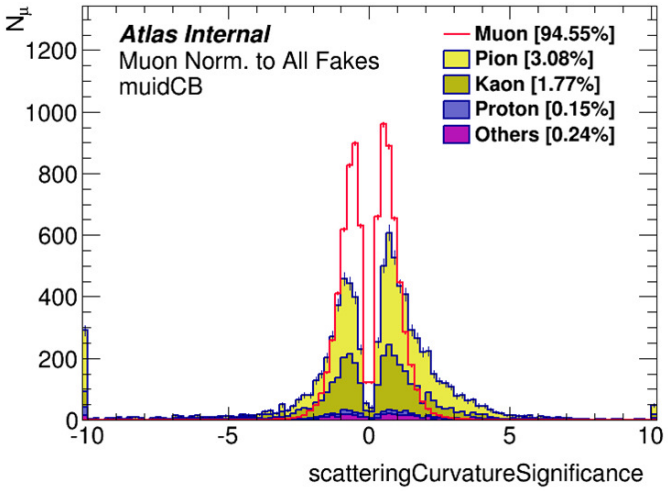


Figure B.4: Comparison of the scattering curvature significance for true muons and various fake muons. All distributions are normalised to the same number.

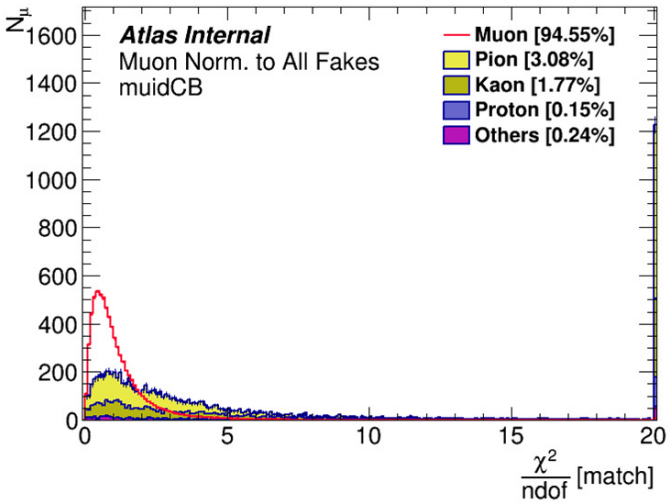


Figure B.5: Comparison of the match quality for true muons and various fake muons. All distributions are normalised to the same number.

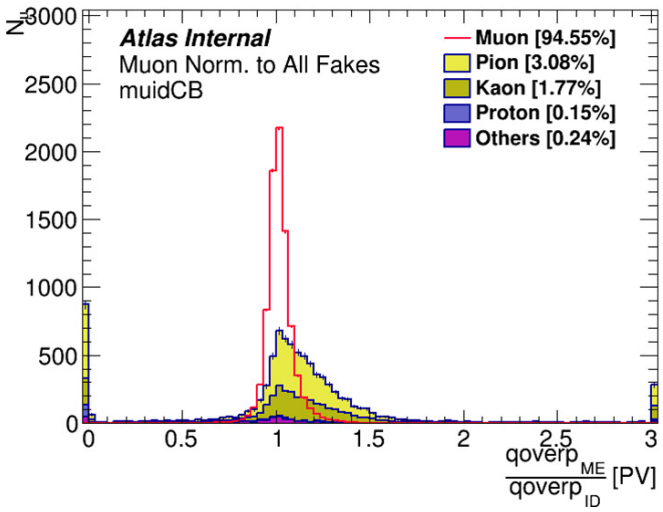


Figure B.6: Comparison of the qoverp ratio between MS and ID tracks for true muons and various fake muons. All distributions are normalised to the same number.

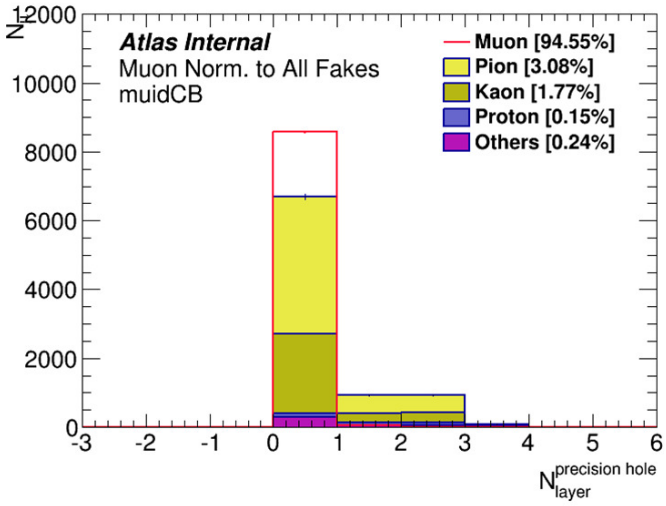


Figure B.7: Comparison of the number of precision holes for true muons and various fake muons. All distributions are normalised to the same number.

Table B.1: Performance comparison between medium+ and BDT in terms of fake reduction at the same signal efficiency. The numbers in the first and second rows are true muon efficiency or fake muon reduction rate. The numbers in the last row are the relative difference between the two algorithms.

Type	Muon	Pion	Kaon	Proton	Others	Unknown
Medium+	97.87	-37.22	−37.02	−60.07	−55.74	−28.35
BDT	97.83	−59.60	−57.25	−80.78	−82.64	−37.12
Improvement	−0.04	60.12	54.66	34.49	48.25	30.92

The true muon sample and π/K samples are fed into the BDT training in terms of the above variables. The resulting shape of the BDT as well as its performance are shown in Fig. B.8. From the receiver operating characteristic (ROC) curve, we can see that when choosing 90% signal efficiency working point, the BDT rejects 80% fakes from π or K decays.

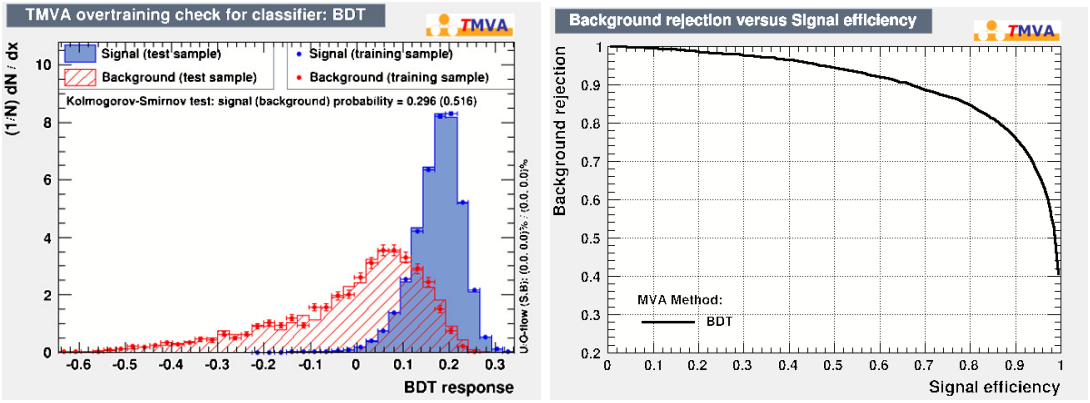


Figure B.8: Overtraining check of the BDT and its ROC curve illustrating its performance.

The performance is also checked by applying this BDT to the muons in the above $t\bar{t}$ sample and compared with the so-called "medium+" cuts provided by MCP which is a generally applied selection by various physical analyses. To have a fair comparison the BDT working point is chosen at which we have about the same true muon efficiency as the medium+ has. The results are shown in Table B.1. As can be seen, the fake muon reduction is largely improved with the BDT method with respect to the medium+, while they have almost the same true muon efficiency. Another conclusion is that, the pion and kaon always have similar fake reduction which is expected since the reasons they fake a muon are the same.

B.3 Sample dependence

Since we are using a special $t\bar{t}$ sample to train the BDT, it may not be applicable to muons in other environment. To assess the potential bias, a $Z \to \mu\mu$ sample is used to check the sample dependence of the input variables for the BDT. The selection procedure is exactly the same as those described above. The results are shown in Fig. B.9. From the comparison, the fake muon samples show no significant sample dependence. While for the true muons, the momentum balance significance, the scattering curvature significance, and the Qoverp ratio clearly are sample-dependent.

One possible explanation is that some of the variables could be p_T dependent since the true muons from the hadronic decays of the $t\bar{t}$ sample are mostly non-isolated muons which tend to be softer in p_T while those in the $Z \to \mu\mu$ sample are prompt muons which tend to be harder. As a first try, we reweight the above problematic variable distributions of the $Z \to \mu\mu$ sample to the $t\bar{t}$ sample according to the true muon p_T distribution. The results are shown in Fig. B.10 indicating the explanation is correct. To further check this, the BDT output distributions are compared between the two samples before and after the p_T reweighting as shown in Fig. B.11. The disagreement in BDT shape of true muons between the samples are cured by the reweighting. Thus we perform a p_T binned BDT training. A preliminary result is shown in Fig. B.12. Some observations are in order: the low performance in the $[4,\ 8]\,\mathrm{GeV}$ p_T bin which can be seen from the smaller area of the corresponding ROC curve (this might be improved by using $J/\Psi \to \mu\mu$ samples) and the low statistics in the high p_T bins as can be seen from the obvious fluctuations of the curves (this might be fixed by merging the two or three high p_T bins).

B.4 Variable validation

So far, the study is based on the MC samples. For the same technique to be applied in real data, we need to compare the input distributions between data and MC samples. This is done by selecting a $Z \to \mu\mu$ control region by requiring the invariant mass of two muons be in a Z mass window. Then the two muons that lead to the Z mass reconstruction provide "true" muon samples with high purity, while the remaining muons in the event if any are considered as "fake" one which has a purity of $\sim 50\%$. The results are shown in Fig. B.13 where the yellow band includes

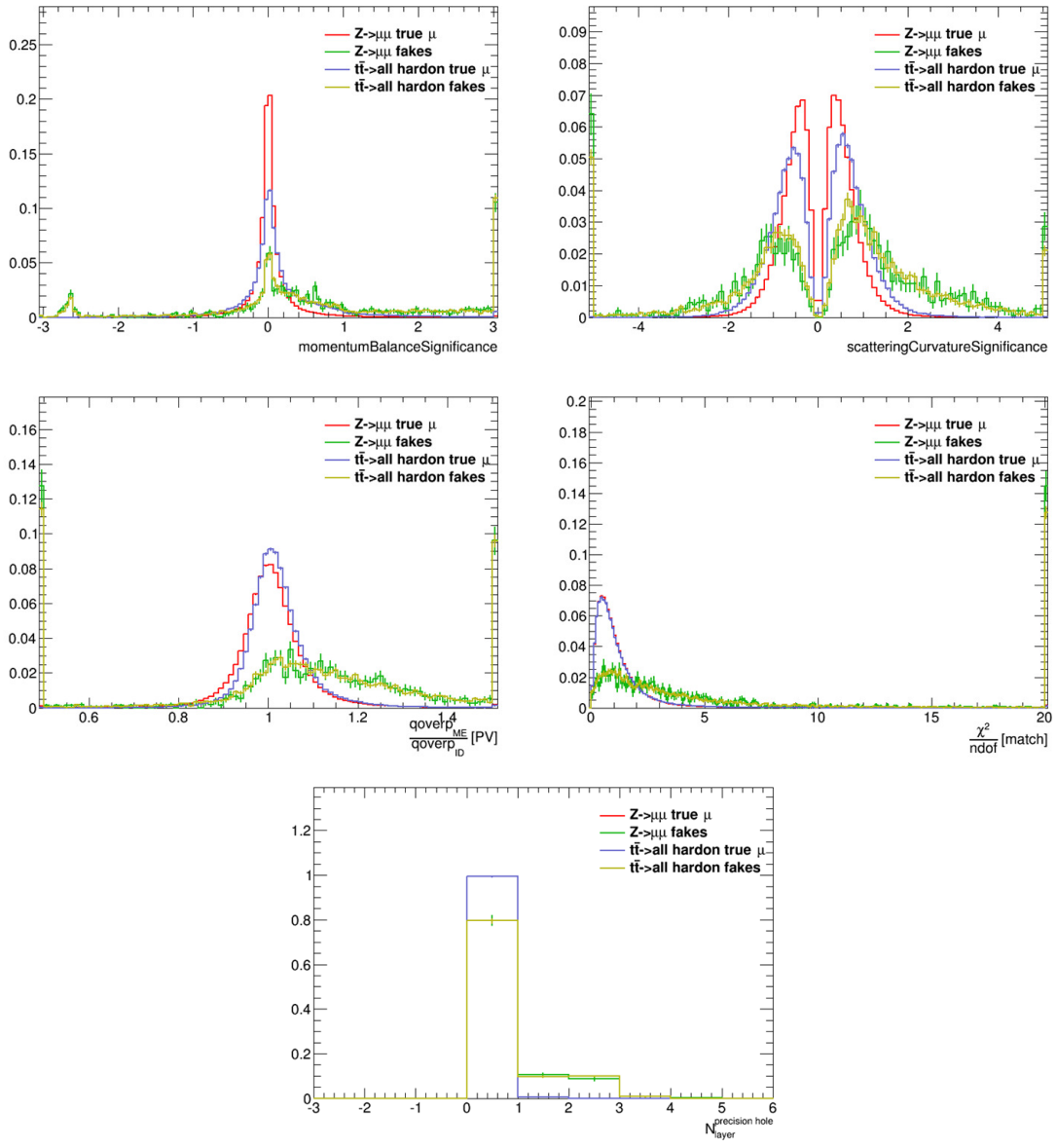


Figure B.9: Sample dependence check of the true/fake muon distributions used for BDT training.

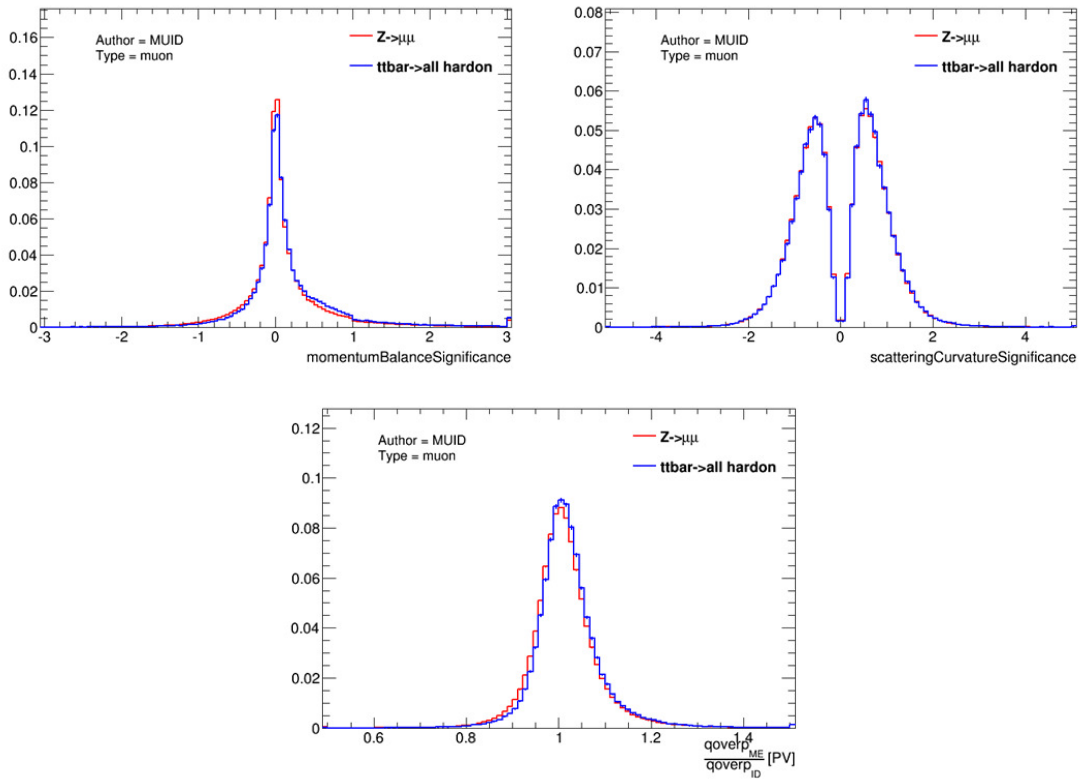


Figure B.10: Sample dependence check of the true/fake muon distributions used for BDT training after p_T reweighting.

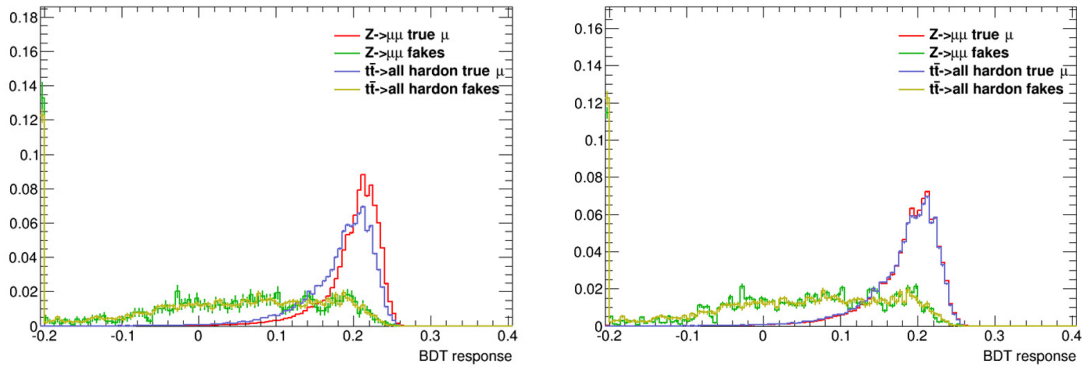


Figure B.11: Sample dependence check of the true/fake muon distributions used for BDT training after p_T reweighting.

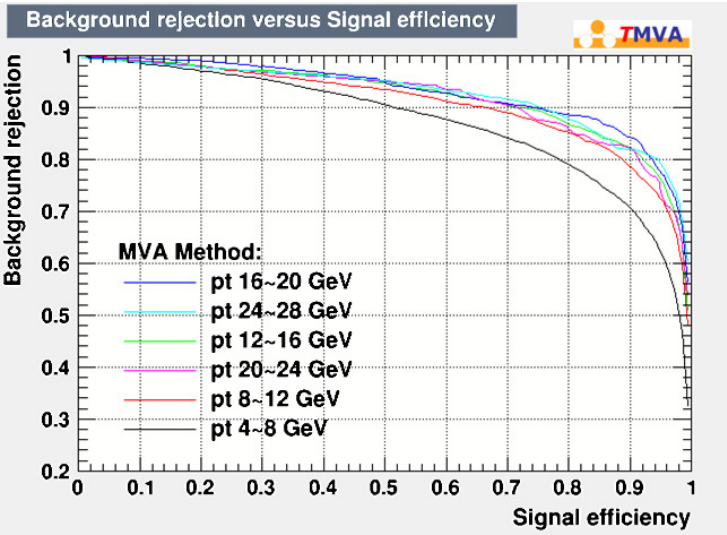


Figure B.12: Performance comparison of BDTs trained in each muon p_T bins in terms of ROC curves.

both statistical and systematic uncertainties. The muon's kinematics (p_T and η) are well modelled. But other variables are not. However, the deviations lie mostly in the lower statistics tails of the distributions which may indicate possible missing processes there since we have not included the W+jets and QCD contributions in the figures yet. The data/MC difference needs further investigation.

B.5 Summary

A preliminary study for the fake muon suppression in the muon reconstruction of the 3rd muon chain for Run II has been performed using the BDT technique trained from and tested on $t\bar{t} \to hh$ and $Z \to \mu\mu$ samples. Several promising discriminating variables are identified. The output BDT shows good performance. This same technique is now being used in the $B_{(s)} \to \mu\mu$ analysis [107]. Further studies are needed to validate the application of the method in data with these variables.

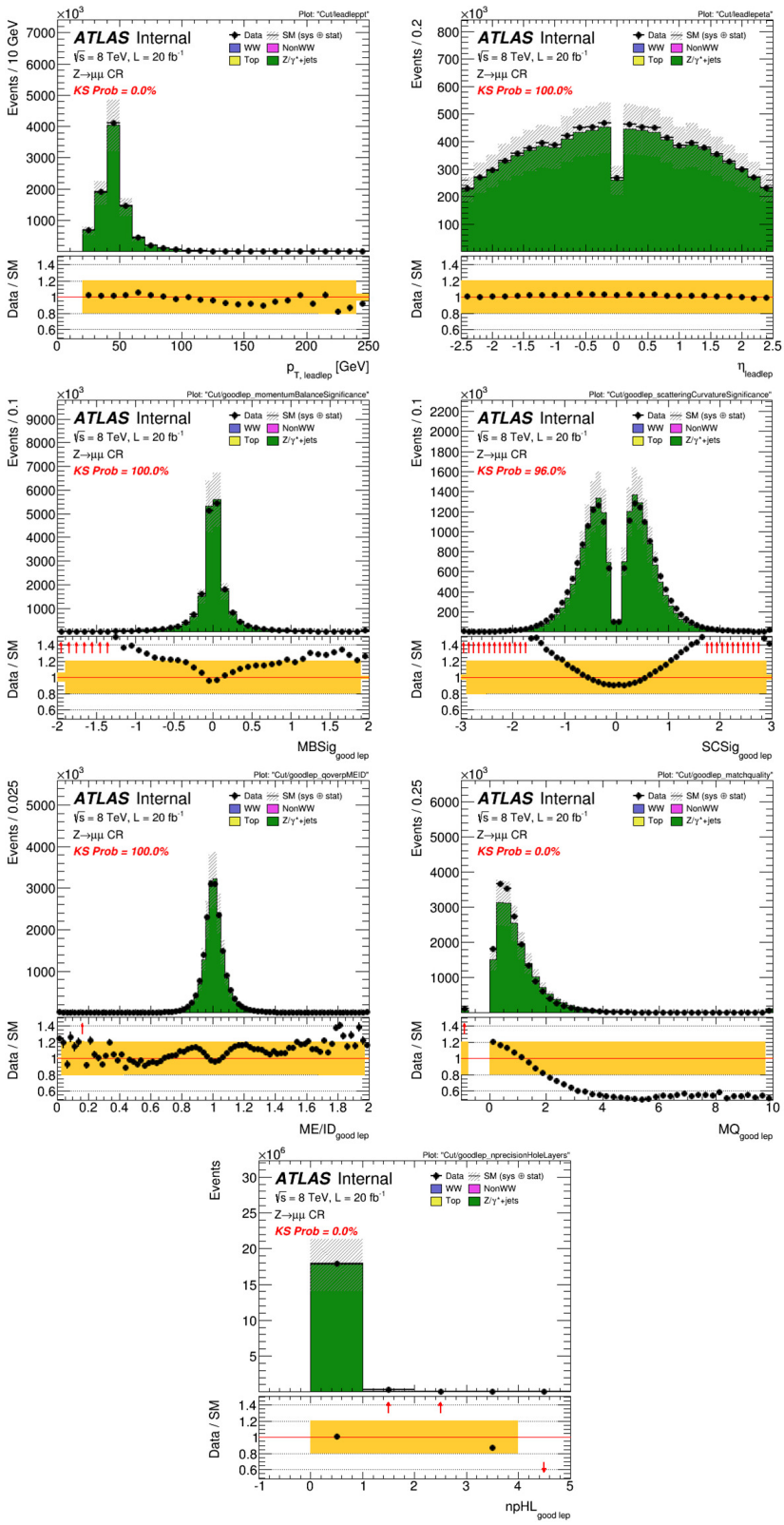


Figure B.13: True muon shape checks in the Z peak region using $Z \to \mu\mu$ sample.

Bibliography

[1] ATLAS Collaboration (G. Aad et al.), Phys. Lett. B **716** (2012) 1–29

[2] CMS Collaboration (S. Chatrchyan et al.), Phys. Lett. B **716** (2012) 30–61

[3] ATLAS Collaboration (G. Aad et al.), Phys. Lett. B **726** (2013) 88–119

[4] ATLAS Collaboration (G. Aad et al.), Phys. Lett. B **726** (2013) 120–144

[5] M. Maggiore, A Modern introduction to quantum field theory (Oxford, 2005)

[6] M. Robinson, Symmetry and the Standard Model: Mathematics and Particle Physics (Springer, 2011)

[7] ATLAS Collaboration (G. Aad et al.), Phys. Rev. D **90** (2014) 052004

[8] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321

[9] P.W. Higgs, Phys. Lett. **12** (1964) 132

[10] P.W. Higgs, Phys. Rev. Lett. **13** (1964) 508

[11] G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, Phys. Rev. Lett. **13** (1964) 585

[12] L.N. Lipatov, Sov. J. Nucl. Phys. **20** (1975) 94

[13] V.N. Gribov, L.N. Lipatov, Sov. J. Nucl. Phys. **15** (1972) 438

[14] G. Altarelli, G. Parisi, Nucl. Phys. B **126** (1977) 298

[15] L. Yu Dokshitzer, Sov. Phys. JETP **46** (1977) 641

[16] S. Alekhin, J. Blumlein, S. Klein, and S. Moch, Phys. Rev. D **81** (2010) 014032

[17] P. Nadolsky et al., Phys. Rev. D **78** (2008) 013004

[18] H1 Collaboration (F.D. Aaron et al.), JHEP **1001** (2010) 109

[19] V. Radescu for the H1 and ZEUS Collaborations, arXiv:1107.4193 [hep-ex] (2011)

[20] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt, Eur. Phys. J. C **63** (2009) 189

[21] R.D. Ball et al., Nucl. Phys. B **867** (2013) 244–289

[22] K. Jakobs, arXiv:1206.7024 [hep-ex] (2012)

[23] ATLAS Collaboration (A. Airapetian et al.), ATLAS: Detector and physics performance technical design report, ATLAS TDR, CERN, Geneva, May 1999, CERN-LHCC-99-14

[24] P. Loch for the ATLAS Collaboration, J. Phys.: Conf. Ser. **323** (2011) 012002

[25] ATLAS Collaboration, Expected electron performance in the ATLAS experiment, ATL-PHYS-PUB-2011-006, CERN, Geneva, Apr 2011

[26] ATLAS Collaboration, Expected photon performance in the ATLAS experiment, ATL-PHYS-PUB-2011-007, CERN, Geneva, Apr 2011

[27] ATLAS Collaboration, Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data, ATLAS-CONF-2014-032, CERN, Geneva, Jun 2014

[28] M. Cacciari, G.P. Salam and G. Soyez, FastJet, http://fastjet.fr/

[29] G.P. Salam and G. Soyez, JHEP **0705** (2007) 086

[30] S. Frixione and B.R. Webber, JHEP **0206** (2002) 029

[31] G. Corcella et al., JHEP **0101** (2001) 010

[32] P. Nason, JHEP **0411** (2004) 040

[33] S. Frixione, P. Nason, and C. Oleari, JHEP **0711** (2007) 070

[34] S. Alioli, P. Nason, C. Oleari, and E. Re, JHEP **0206** (2010) 043

[35] J. Allison et al., IEEE Trans. on Nucl. Sci. **53** (2006) 270–278

[36] S. Agostinelli et al., Nucl. Instr. and Meth. A **506** (2003) 250 – 303

[37] ATLAS HWW Working Group, The analysis of $H \to WW \to \ell\nu\ell\nu$ ggF and VBF production modes with 20 fb^{-1} and 5 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s} = 8$ and 7 TeV, ATL-COM-PHYS-2014-466, CERN, Geneva, May 2014

[38] ATLAS Collaboration (G. Aad et al.), arXiv:1503.03643 [hep-ex] (2015) , submitted to Eur. Phys. J. C

[39] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, JHEP **1202** (2012) 88

[40] J. Alwall et al., JHEP **1407** (2014) 079

[41] S. Alioli, P. Nason, C. Oleari, and E. Re, JHEP **0904** (2009) 002

[42] T. Sjostrand, S. Mrenna, and P. Z. Skands, Comput. Phys. Commun. **178** (2008) 852–867

[43] P. Nason and C. Oleari, JHEP **1002** (2010) 037

[44] J. Alwall et al., JHEP **0709** (2007) 028

[45] T. Binoth, M. Ciccolini, N. Kauer, and M. Kramer, JHEP **0612** (2006) 046

[46] S. Frixione and B. Webber, JHEP **0308** (2003) 007

[47] B. P. Kersevan and E. Richter-Was, Comput. Phys. Commun. **184** (2013) 919–985

[48] M.L. Mangano et al., JHEP **0307** (2003) 001

[49] ATLAS Collaboration (G. Aad et al.), arXiv:1412.2641 [hep-ex] (2014) , Submitted to Phys. Rev. D

[50] A. Justel, D. Peña, and R. Zamar, Stat. and Prob. Lett. **35** (1997) 251 – 259

[51] R.Keith Ellis, I. Hinchliffe, M. Soldate, and J.J. van der Bij, Nucl. Phys. B **297** (1988) 221

[52] ATLAS HWW working group, Search for the Higgs Boson in Association with a Weak Boson and Decaying to a W-boson Pair with the ATLAS Detector at the LHC, ATL-COM-PHYS-2013-1560, CERN, Geneva, November 2013

[53] Y. Freund and R.E. Schapire, J. of Comput. and Syst. Sci. **55** (1997) 119 – 139

[54] J.M. Campbell, R.Keith Ellis, and C. Williams, JHEP **1107** (2011) 018

[55] ATLAS HWW Working Group, Background estimation in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis with 20.7 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s} = 8$ TeV, ATL-COM-PHYS-2013-1630, CERN, Geneva, Dec 2013

[56] H.-C. Cheng and Z. Han, JHEP **0812** (2008) 063

[57] B. Mellado, X. Ruan, and Z. Zhang, Phys. Rev. D **84** (2011) 096005

[58] ATLAS HWW Working Group, Background estimates in the $H \to WW^* \to \ell\nu\ell\nu$ analysis with 21 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s} = 8$ TeV, ATL-COM-PHYS-2013-139, CERN, Geneva, Feb 2012

[59] ATLAS HWW Working Group, Update of the $H \to WW^* \to \ell\nu\ell\nu$ analysis with 21 fb^{-1} of $\sqrt{s} = 8$ TeV data and 4.6 fb^{-1} of $\sqrt{s} = 7$ TeV data collected with the ATLAS Detector, ATL-COM-PHYS-2013-159, CERN, Geneva, Feb 2013

[60] ATLAS HWW Working Group, Analysis of $H \to WW^* \to \ell\nu\ell\nu$ in the same-flavour channels, ATL-COM-PHYS-2012-1445, CERN, Geneva, Sep 2012

[61] The LHC Higgs Cross Section Working Group (S. Heinemeyer et al.), ArXiv 1307.1347 (2013)

[62] A. Bredenstein, A. Denner, S. Dittmaier, and M.M. Weber, Phys. Rev. D **74** (2006) 013004

[63] C. Anastasiou and K. Melnikov, Nucl. Phys. B **646** (2002) 220

[64] M. Spira, A. Djouadi, A. D. Graudenz, and P.M. Zerwas, Nucl. Phys. B **453** (1995) 17

[65] S. Catani and D. de Florian and M. Grazzini, and P. Nason, JHEP **0307** (2003) 028

[66] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini, Phys. Lett. B **595** (2004) 432

[67] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, Phys. Lett. B **670** (2008) 12

[68] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, JHEP **1111** (2011) 064

[69] K. Hamilton, P. Nason, and G. Zanderighi, JHEP **1210** (2012) 155

[70] J. Butterworth et al., arXiv:1405.1067 [hep-ph] (2014)

[71] A. Banfi, G. Salam, and G. Zanderighi, JHEP **1206** (2012) 159

[72] I. Stewart and F. Tackmann, Phys. Rev. D **85** (2012) 034011

[73] P. Bolzoni, F. Maltoni, S. Moch, and M. Zaro, Phys. Rev. Lett. **105** (2010) 011801

[74] M. Ciccolini, A. Denner, and S. Dittmaier, Phys. Rev. D **77** (2008) 013002

[75] J. Pumplin, J. Huston, H.L. Lai, Wu-Ki Tung, and C.-P. Yuan, Phys. Rev. D **80** (2009) 014019

[76] A. Bierweiler, T. Kasprzik, Johann H. Kühn, JHEP **1312** (2013) 071

[77] PDF4LHC Working Group (S. Alekhin et al.), arXiv:1101.0536 [hep-ph] (2011)

[78] S. Frixione, E. Laenen, P. Motylinski, B.R. Webber, and C. White, JHEP **0807** (2008) 029

[79] ATLAS HWW Working Group, Theoretical studies for the $H \to WW$ measurement with the ATLAS detector, ATL-COM-PHYS-2013-1541, CERN, Geneva, Nov 2013

[80] ATLAS Collaboration (G. Aad et al.), Eur. Phys. J. C **73** (2013) 2518

[81] ATLAS Collaboration (G. Aad et al.), Eur. Phys. J. C **71** (2011) 1630

[82] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. C **71** (2011) 1554

[83] ATLAS Collaboration (G. Aad et al.), Phys. Rev. D **90** (2014) 052004

[84] P. Artoisenet et al., JHEP **1311** (2013) 043

[85] L. Randall, R. Sundrum, Phys. Rev. Lett. **83** (1999) 3370–3373

[86] ATLAS HWW Spin Working Group, Spin/Parity $H \to WW^* \to \ell\nu\ell\nu$ analysis with $20\,\text{fb}^{-1}$ of data collected with the ATLAS detector at 8 TeV, ATL-COM-PHYS-2014-264, CERN, Geneva, Apr 2014, Supporting note for the spin paper

[87] A. Read, J. Phys. G **28** (2002) 2693–2704

[88] F. Caola and K. Melnikov, Phys. Rev. D **88** (2013) 054024

[89] J. Campbell, R.Keith Ellis, and C. Williams, JHEP **1404** (2014) 060

[90] N. Kauer and G. Passarino, JHEP **1208** (2012) 116

[91] O. Arnaez, J. Long, Z. Zhang, Y. Huang, Y. Li, and M. Duehrssen, Constraints on the off-shell Higgs coupling and the Higgs boson width in the $H \to WW$ channel, ATL-COM-PHYS-2014-1322, CERN, Geneva, Oct 2014

[92] M. Bonvini, F. Caola, S. Forte, K. Melnikov, and G. Ridolfi, Phys. Rev. D **88** (2013) 034032

[93] ATLAS Collaboration (G. Aad et al.), arXiv:1503.01060 [hep-ex] (2015) , Submitted to Eur. Phys. J. C

[94] Y. Li, Y X. Ruan, L. Yuan, and Z. Zhang, Review of data-driven methods used to control the normalization of the top quark background contribution in the $H \to WW^* \to \ell\nu\ell\nu$ analyses at the LHC, ATL-PHYS-INT-2014-037, CERN, Geneva, Dec 2014, LAL 13-373

[95] ATLAS Collaboration (G. Aad et al.), Phys. Lett. B **712** (2012) 289–308

[96] ATLAS Collaboration (G. Aad et al.), Phys. Lett. B **716** (2012) 62–81

[97] A. Belloni, G. Conti, J. Guimaraes da Costa, J. Huth, and C. Mills, Top Background Estimation Using Jet Multiplicity Distributions, ATL-COM-PHYS-2012-856, CERN, Geneva, Jun 2012

[98] ATLAS Collaboration (G. Aad et al.), Phys. Rev. D **87** (2013) 112001

[99] CMS Collaboration (S. Chatrchyan et al.), Phys. Lett. B **710** (2012) 91–113

[100] CMS Collaboration (S. Chatrchyan et al.), Eur. Phys. J. C **73** (2013) 2469

[101] ATLAS Collaboration (G. Aad et al.), Phys. Lett. B **718** (2013) 879

[102] M. Bohler, DESY-THESIS-2012-022, CERN-THESIS-2012-245 (2012)

[103] ATLAS Collaboration (G. Aad et al.), Eur. Phys. J. C **72** (2012) 2056

[104] CMS Collaboration (S. Chatrchyan et al.), Phys. Rev. Lett. **109** (2012) 261802

[105] S. Hanany, M.D. Niemack, and L. Page, CMB Telescopes and Optical Systems, p. 431, in *Planets, Stars and Stellar Systems. Volume 1: Telescopes and Instrumentation*, edited by T.D. Oswalt and I.S. McLean, (, 2013)

[106] ATLAS Collaboration (G. Aad et al.), Phys. Lett. B **718** (2013) 860–878

[107] C. Alpigiani et al., Analysis for $B^0_{(s)} \to \mu\mu$ decays with $20\,\text{fb}^{-1}$ of integrated luminosity, ATL-COM-PHYS-2014-1179, CERN, Geneva, Sep 2014.

Appendix C

Publications

The three ATLAS papers, one top review article and four supporting notes in which I have contributed are listed below:

1. *Observation and measurement of Higgs boson decays to WW^* with ATLAS at the LHC* [49],

2. *Determination of spin and parity properties of the Higgs boson in the $WW^* \to e\nu\mu\nu$ decay channel* [38],

3. *Determination of the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector* [93],

4. *Review of data-driven methods used to control the normalisation of the top quark background contribution in the $H \to WW^* \to \ell\nu\ell\nu$ analyses at the LHC* [94].

5. *Theoretical studies for the $H \to WW$ measurement with the ATLAS detector* [79],

6. *Background estimation in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis with 20.7 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s}$ = 8 TeV* [55],

7. *Spin/Parity $H \to WW^* \to \ell\nu\ell\nu$ analysis with 20 fb^{-1} of data collected with the ATLAS detector at 8 TeV* [86]

8. *Constraints on the off-shell Higgs coupling and the Higgs boson width in the $H \to WW$ channel* [91].

Appendix D

Acknowledgments

First of all, I'd like to thank my co-supervisor in France, Dr. Zhiqing Zhang, who guided me into the world of high energy physics (HEP) step by step in our daily discussion at L.A.L. And without his pushing forward on me, I will not have finished my thesis advanced in time. He also gave me precious suggestions on my thesis writing and polished my English word by word, which I appreciated very much. Besides, he offered me a lot of help during my stay in France especially when I couldn't read those papers/letters written in French. By the way, I want to thank the Chinese Scholarship Council (CSC), without whose support I would never have the chance to work with Zhiqing.

Then I want to thank my supervisor at Nanjing University, Professor Shenjian Chen, who accepted me as his PhD student when I was wondering where to go just after I finished my undergraduate years. He was also our QM teacher in my undergraduate. Due to his excellent teaching that time, I was inspired to join his research group. He provided me the chance to do my research about the Higgs in ATLAS. He is a very gentle man and always smile on you which makes you feel easy or sometimes over-easy.

There are other teachers I want to give thank for: Dr. Shan Jin in IHEP, I stayed under his supervision for half an year, he impressed me a lot for his pushing on work and his humour in words; Dr. Lianyou Shan in IHEP, we did not meet too much, but every time we sit and talked, he gave me advise and courage for continuing on the way of HEP; Dr. Weimin Sun, he was our QFT teacher, he solved me puzzles many times.

Due to my limited French, I did not communicate so frequently with people at LAL. But I still have a number of people to thank, especially: RD, the former

ATLAS group leader, who speaks fluent English, he is amiable and talked to me ofter; Dirk, who guided my first tour to CERN; Louis and his wife Lydia, we "salut" to each other a lot, especially in weekends and they are really workaholics to me; Geneviève and Sylvie, who were the secretaries, they were patient and warm-hearted; Isabelle, the responsible person in the "service missions", who helped to deal with the allowance forms.

I'd like to thank people in the HWW group: Lina, the contact person of the low p_T sub-group that I firstly involved in; Tomoe, who was in charge of the sample production, he never disappointed me when I had question on the common ntuples (a sample structure used in our group); Jonathan, who was one of the core developers of the CAF (Common Analysis Framework), he helped me many times when I got into trouble with the code and recently we worked together on the Higgs width analysis, especially on cross checking the cutflow tables; Tatjana, the contact person of the spin sub-group, she offered me the chance to take part in the Higgs spin analysis; Doug, who shared his code with me generously when I worked in the spin group.

I also have some people to thank outside the HWW group: Elmar, who introduced to me the basic idea of the ISF; Niels and Massimo, the conveners in the MCP group, they discussed with me often when I was at CERN on my qualification task on the fake muon study.

Besides, I'd like to thank Achille Stocchi, the director of LAL and all the jury members of my thesis defence: Yeuk-kwan Edna Cheung, Cristinel Diaconu, Yuanning Gao, Feng Liu, Yajun Mao, and Haijun Yang.

In addition, I also have a lot of friends to thank. They are all PhD students or postdocs in the ATLAS experiment. I'd like to thank: Xifeng Ruan who gave me basic training on data analysis in the HWW group; Li Yuan, we cooperated a lot inside the HWW group, Lei Zhang, we discussed practical questions in many aspects and he, together with Xifeng and Li, often offered me dormitory when CERN hotel was unavailable; Jie Yu, who was in the top group so I seeked for advise when I met problems on top quark events; Yanping, we met already in IHEP and cooperated closely recently.

Then I'd like to thank my girlfriend Wanwan Li. Thanks for her understanding of my job and for her support during my thesis writing.

At last, I want to give my highest appreciation to my parents. They are always so kind to me and make me feel warm when things are tough. Although they don't understand my job, they support me with all the best they can do.